Genetic Code Domains Conserve the Imprint of tRNA Cofactors Encoded to Specify Cognate Amino Acid Synthesis[1]

Brian K. Davis

Research Foundation of Southern California, Inc.
8837 Villa La Jolla Dr. #13595, La Jolla, CA 92037

(*preprint date*: August 22, 2011)

Proteins are assembled almost exclusively from a set of only twenty amino acids, in five synthesis families. Same-family amino acids generally charge related pre-species-divergence tRNA that read codons within a contiguous region of the genetic code, termed a domain. tRNA sets formed using an arbitrary pairing rule, illustrated by matching codon-cognate and anticodon-cognate tRNA or by randomly regrouping these tRNA, mainly pair unrelated tRNA from different code domains. They display reduced N2-base identity (elevated N2-base complementarity) at this GC rich site, low pre-divergence sequence identity, low anticodon contiguity, and negatively correlated amino acid path-distances, in contrast to same-domain pairs. Sibling amino acid reliance on related tRNA, cognate with contiguous codons, links genetic code formation with the growth of nascent tRNA-dependent amino acid synthesis pathways. Path-identity elements in these cofactor/adaptor tRNA molecules furnish a pre-synthetase, RNA-based mechanism for specifically pairing amino acids and their codons.

[1] A preliminary report of the present findings can be found at: http://www.archive.org/details/ModularStructureOfTheGeneticCodeProvidesALinkToTheCompleteRnaCode

Significant progress has been made in clarifying the structure of the genetic code since its nearly universal set of codon assignments was uncovered in the mid-1960s[1]. A diverse set of structural regularities, identified by several investigators, is now attributed to the code[2] - an annotated list of thirty-two code features appears in Table A1. As their number and scope grew, the 'frozen accident' scenario[3], suggesting codon assignments were purely random, became increasingly implausible. It appeared more likely instead that each feature identified portrayed a different aspect of a deep structure within the code[4-7]. Consistent with this, a model of code formation that equated the time-order of amino acid addition to the code with the number of reaction steps required for synthesis (Fig. S1) succeeded in unifying many features of code structure[2,5]. It became apparent, for example, why NAN and NUN triplets specifically code for hydrophilic and hydrophobic residues respectively[8], when any combination of mid-base, or 5'-base, codon sets, in principle, could serve to optimize amino acid homology[9]. 5'-Base invariance among codons for same-family amino acids and assignment of a codon set-of-four (3'-base degenerate) to each of six small amino acids[10] were among other puzzling features of code structure elucidated by the path-distance model of code formation[2,5]. The proposed model will be shown here to also explain N2-base complementarity, produced by counter-aligned acceptor stem N2:N71 bp, in tRNA species with complementary anticodons[11]. This correlation has been offered as evidence that the acceptor stem 5'-triplet, centered at N2, encoded amino acids before the anticodon (N34-N36). With identity elements in the tRNA acceptor stem[12-14], N2-base complementarity additionally offered the possibility of providing an insight into amino acid recognition before appearance of aminoacyl-tRNA synthetase (aaRS) enzymes. A non-protein amino acid recognition mechanism of some kind existed in the pre-aaRS era. While it remains unknown, the nature of the initial RNA code, apart from a set of monolingual RNA triplet interactions, shall be an open question[15].

Recent investigations on code structure[2,4-7,16-19] reveal that depicting the genetic code solely as a set of amino acid/codon pairs[1,3] understates the complexity of code organization: (i) Codon mid-base,

set size, and distribution correlations with amino acid path-distances are omitted[2,5,6]. And, (ii) partitioning of the code into domains of biosynthetically related amino acids with phylogenetically related pre-divergence (post-divergence sequence variations[20] filtered-out) tRNA species, possessing the same core structure group[2,21] and cognate with nearest-neighbor codons, is also omitted[7,19]. Fifteen amino acids are distributed among five code domains. Five remaining amino acids, from the standard set, are in three small quasidomains, which complete the modular organization of the code. These findings provide the strongest evidence yet obtained that bifunctional tRNA[22,23], acting as cofactors in amino acid synthesis and as adaptors in translation, coordinated code expansion with the growth of amino acid synthesis pathways[2,7]. Asparagine and glutamine[24,25] and a few other amino acids[26-28] retain this mode of synthesis, principally among prokaryote species. Code domains conserve the imprint of a once extensive network of tRNA-dependent amino acid synthesis pathways[7,19]. As they span regions of contiguous codons[2,7,19], any arbitrary tRNA set with respect to code structure, such as that obtained by pairing codon-cognate and anticodon-cognate tRNA species[11], should generally match tRNA whose anticodons differ by more than one base. Sequence heterogeneity between unrelated, mixed-domain tRNA species could be anticipated, therefore, to contribute to elevated N2-base complementarity among tRNA pairs with complementary anticodons. A survey of base identity in the trace of pre-divergence tRNA species conserved across sources in Archaea, Bacteria and Eukarya has confirmed this. Evidence of a reliance on tRNA cofactors in amino acid synthesis during code formation, furthermore, furnishes an RNA mechanism for encoding amino acids operative in the pre-aaRS era.

Methods

tRNA Sets

Twenty-four tRNA pairs with complementary anticodons, drawn from Rodin et al.[11], formed the primary set in this investigation. Three other tRNA sets, with distinct pairing rules, have been

examined that were produced by regrouping complementary set tRNA species. Canonical bp interactions between anticodons in the complementary set uniquely defined tRNA pairs. The consensus base among the phylogenetically determined ancestral tRNA, in the three species Kingdoms, determined the N2-base of each pre-divergence tRNA species. Nine pairs in the complementary set involved a eukaryote tRNA bearing an A34. For three of these pairs, a eubacterial tRNA with an A34 occurred also, and in two pairs it shared the N2-base of the ancestral eukaryote adaptor. In four pairs, eukaryote tRNA-A34 shared the same N2-base as a eubacterial and/or archaeal isoacceptor having a G34 (transition variant). The N2-base could not be inferred for pre-divergence tRNA in two other pairs, owing to the rarety of prokaryote tRNA-A34. All complementary set pairs with contiguous anticodons possessed counter-aligned N2:N71 bp (Table A2). On this basis, the set was subdivided into subsets with contiguous and non-contiguous anticodons. A third tRNA set was formed by rearranging complementary set tRNA into twenty-two same-domain (same-quasidomain) pairs. Four tRNA from different code domains were left unpaired in this procedure. This set also included 4 pairs of isoacceptors cognate with codons in different sets-of-four. Same-domain tRNA species having cognate amino acids with comparable path-distances were generally paired, to minimize differences in adaptor ‘code age’. A fourth set contained pairs of tRNA with randomly matched code domains. One member of each pair in the same-domain set, and unpaired tRNA, was randomly redistributed to produce the pairs of this set.

Quaternary sequence identity in pre-divergence tRNA. The trace of pre-divergence tRNA conserved within extant tRNA sequences was identified and used to assess kinship between tRNA species that predated the Last Common Ancestor (LCA). Kinship between pre-divergence tRNA being determined by the amount of shared information retained within the conserved trace of their base sequence. tRNA sites invariant between consensus sequences obtained from sources in Archaea, Bacteria, and Eukarya were located[7] among 1,100 base sequences of 48 tRNA species found during a search of the tRNA database at: http://trnadb.bioinf.uni-leipzig.de/search (Table A3).

Base identity at jointly conserved sites, excluding sites universally conserved by tRNA molecules, in the conserved trace of pre-divergence base sequences is expressed in quaternary units (quarts) as,

$I_{ij} = - \log_4 p(x_{ij})$

where tRNA species i and j have identity, I_{ij}, when the observed number of jointly conserved (non-universal) identical sites, x_{ij}, has a probability, $p(x_{ij})$; occurrence of any base at any specified site having a random probability of ¼. The form of this equation implies that an identity of n quarts corresponds (within a multiplicative constant) to the amount of uncertainty[29] removed, when a test sequence of length n shows identity with a reference sequence of the same length: $I = - \log_4 4^{-n} = n$ quarts. Comparisons between mean quaternary identities for pre-divergence tRNA were performed using a t-test, or OWAV and Tukey's multiple comparison test. The utility of equating identity in the conserved trace of pre-divergence sequences with shared sequence information has been previously demonstrated[7]. Quaternary identity decreased linearly with Jukes-Cantor mutation distance (*d*), from post-species-divergence phylogenetics, $I = 8.9 – 8.5d$ ($R^2 = 0.60$), for tRNA sequences within an interval, $0 \leq d \leq 1$, $0 \leq I \leq 10$, containing 77 per cent of the sequences examined. Since the conserved trace of an ancestral tRNA species, in extant tRNA populations, is likely to account for less than the total sequence, adopting standard phylogenetic procedures to analyze pre-species-divergence evolution of tRNA paralogs will be burdened accordingly by post-divergence base substitutions - the source of most tRNA sequence variability[20]. In the present procedure[7], post-species divergence variations were filtered-out by restricting the analysis of pre-species-divergence base sequences to non-universal sites, jointly conserved by ancestral tRNA species of Archaea, Bacteria, and Eukarya.

Source domains of tRNA pairs. Differences in the frequency of same-domain and mixed-domain tRNA pairs were noted for each pairing rule considered. As different codon/amino acid/tRNA combinations characterize each code domain and quasidomain[5,7,19], the frequency of each pair was

indicative of the relevance of pair-formation criteria to formation of the code. Contingency tables of pairs, in different tRNA sets, were quantitatively appraised using Fisher's exact probability test, or a Chi-square test with Yates' correction for continuity.

Triplet base frequencies. Nucleotide frequencies in acceptor stem terminal triplets and at anticodon sites were determined from aggregate nucleotide (unmodified form) frequencies reported in a survey by Mark and Grosjean[30] of 4096 base sequences in the genbank database (ftp://ncbi.nlm.nih.gov/genbank/ genomes) for tRNA from 13 Archaea, 30 Bacteria and 7 Eukarya species, with known genome sequence. Normalized, cumulative nucleotide distributions for each triplet examined were evaluated for identity and parallel, or anti-parallel, complementarity in goodness of fit comparisons with a Kolmogorov-Smirnov test. Acceptor stem bp frequencies were deduced from canonical pairs between 5'- and 3'-terminal triplets, with any excess bases distributed to wobble partners[16]. A set of normalized bp frequencies indicated a double-helix structure.

N2-base identity and complementarity. The consensus N2-base among ancestral tRNA species for Bacteria, Archaea, and Eukarya was attributed to the pre-divergence antecedent. Phylogenetic analysis of 8,246 sequences determined the tRNA common ancestor in each species Kingdom (based on ref. 11). In four pairs, a eukaryotic tRNA, bearing an A34, shared an N2-base with an archaeal and eubacterial isoacceptor bearing a G34 (transition variant), and the N2-base was attributed to the pre-divergence tRNA species. tRNA pairs with different N2-bases, from counter-aligned N2:N71 bp, produced complementary bases at this G,C rich acceptor stem site. N2-base complementarity served to indicate tRNA heterogeneity. N2-base identity and complementarity among tRNA pairs in different sets were quantitatively evaluated using Fisher's test, or Chi-square distribution.

Distribution of tRNA types and subtypes. Type I tRNA include three, or more, subtypes[7,21] distinguished by distinct D-arm and variable loop bases and base interactions in the core of the tRNA

L-form. Pairs with different types and subtypes of tRNA furnished additional evidence of tRNA heterogeneity, bearing on the relevance of a pairing rule to the formation of the code. Differences in tRNA core structure group distribution between pairs with codon-cognate and anticodon-cognate tRNA versus same-domain tRNA were appraised with Fisher's exact probability test.

Anticodon contiguity. The distribution of contiguous and non-contiguous anticodons among tRNA pairs in a given set provided evidence of heterogeneity arising from pairing tRNA derived from different code domains. This provided another test on the relevance of the criteria used to pair tRNA to code formation. tRNA pairs whose anticodons differed at more than 1 site (non-contiguous) were enumerated and Fisher's test was used to evaluate the significance of differences in anticodon contiguity between tRNA sets.

Correlation between cognate amino acid path-distances. Amino acid path-distance in this investigation corresponds to the number of reaction steps in an amino acid synthesis pathway[31], counted from citrate cycle with path takeovers discounted. The Pearson correlation between path-distances of cognate amino acids in tRNA pairs from each set was calculated and its dependence on pre-divergence tRNA sequence identity evaluated. In this procedure, each tRNA pair was ranked according to its amino acid path-distance: first by the left amino acid and then by the right. A new set, retaining the initial number of pairs, was produced by randomly sampling (with replacement) the rank-pair cumulative frequency distribution, according to the bootstrap method of Efron[32]. On decoding pair-ranks in the new set, left and right amino acid path-distances were recovered and their correlation coefficient calculated. Outlined below (Data Analysis) are subsequent steps in relating the amino acid path-distance correlation coefficient to tRNA identity and establishing its significance, independent of the underlying frequency distribution.

tRNA pairs with mixed-class synthetases. The frequency of pairs having tRNA charged by different classes of aaRS was noted and related to pre-divergence tRNA sequence identity in four

sets. Pair frequencies in each set were evaluated using Fisher's test and regression analysis, to establish the strength of any dependence on pre-divergence tRNA sequence identity. The results could link formation of these pairs to the domain structure of the genetic code. In addition to testing the relevance of a tRNA pairing criteria to the mechanism of code evolution, these observations could furnish evidence on the time of aaRS appearance in relation to code formation.

Data analysis. Four tRNA sets formed with different pairing criteria were quantitatively evaluated for differences in pre-divergence sequence identity, domain distribution, N2-base correspondence, anticodon contiguity, core group frequency, amino acid path-distance correlation, and aaRS class heterogeneity. When two tRNA sets with multiple features were compared, the probabilities for each outcome were combined by the method of Fisher[33] to obtain a global probability for relating both sets.

The frequency of tRNA pairs with mixed-class aaRS, same or complementary N2-bases, contiguous anticodons, and correlated amino acid path-distances was related to sequence identity in pre-divergence tRNA species. 1,000 distributions were generated per tRNA set by the bootstrap method[32], through random sampling (with replacement) of the observed frequency distribution. In this procedure, the distribution of each attribute over sets-of-four (3'-base degenerate) comprising the code was related to the pre-divergence sequence identity among tRNA pairs. Assessment of amino acid path-distance correlation dependence on tRNA identity involved, as noted, random sampling of the observed pairwise path-distance frequency distribution. Each feature of a tRNA set was depicted as a cluster of 250 means. The distribution mean is invariant under this operation, and each set of means formed a normal distribution, in accord with the central limit theorem. Non-overlapping clusters indicated a significant difference between tRNA sets ($P \leq 4 \times 10^{-3}$). 1,000 linear regression equations, with tRNA sequence identity as independent variable, were calculated per set. A significant non-zero dependence on tRNA identity was determined from the 98 per cent (non-parametric) confidence interval about the mean slope, where the average displacement of the 1 and

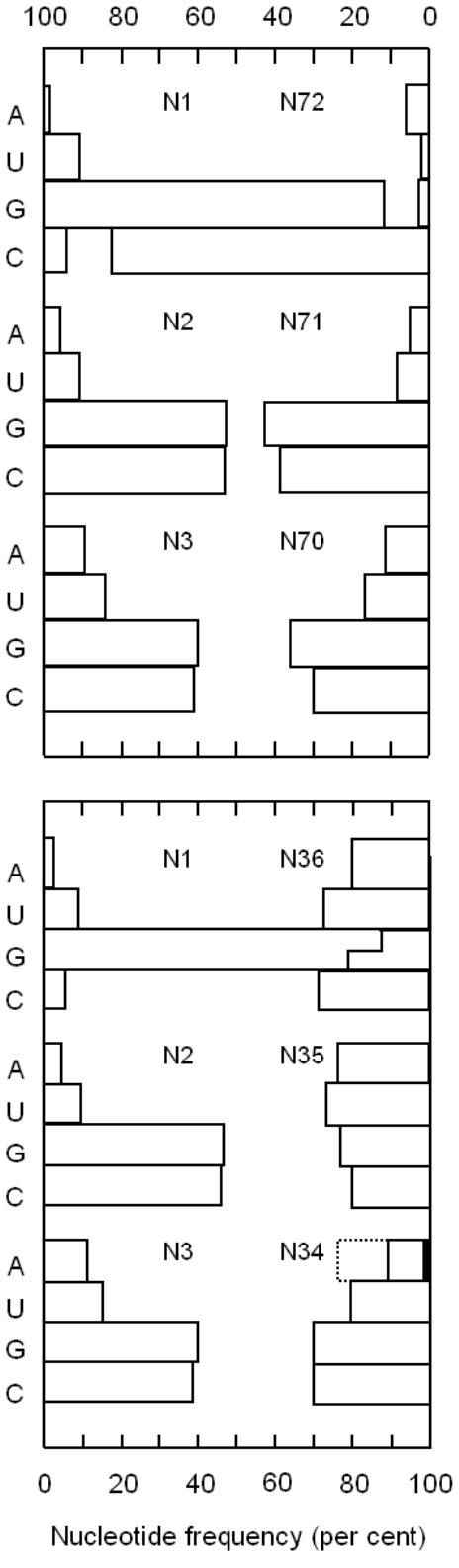
99 percentile among slopes specified the confidence interval. Departures from symmetry[34] within 36 slope distributions produced skewness values within ± ½, and two distributions exceeded this limit: amino acid path-distance correlation, non-contiguous complementary set (0.86) and contiguous complementary set (1.40).

RESULTS

Triplet base distributions

Nucleotide frequencies in acceptor stem terminal triplets and the anticodon reported by Mark and Grosjean[30] for tRNA sequences of 50 species, with known genome sequence, from Bacteria, Archaea, and Eukarya appear in Fig. 1. Complementary (anti-parallel) nucleotide profiles characterize acceptor stem terminal triplet sites N1-N3 and N72-N70 ($\Delta_{max}(n_{\downarrow\uparrow} = 12) = 0.167$, P = 0.991, goodness of fit). Anticodon sites, N34-N36, display comparatively uniform nucleotide profiles, discounting the near exclusion of A at N34 in prokaryote tRNA (Fig. 1a). No detectable identity, or complementarity (parallel, anti-parallel), existed between 5'-acceptor stem triplet, N1 – N3, and anticodon base distributions. Differences between them were almost significant ($\Delta_{max}(n = 12) = 0.50$, P = 0.066), making an acceptor-stem triplet to anticodon transformation[11] an unlikely explanation for a correlation between tRNA N2-base and anticodon complementarity. Each site in the terminal acceptor stem triplet contains a distinct base profile. G1**:**C72 bp occurred in nearly 84 per cent of tRNA sequences (Fig. 1B,C). About half the remaining tRNA (7.1 per cent) contained a U1**:**A72 bp and roughly a quarter had a C1**:**G72. Wobble pair, G**:**U, represented 2.1 per cent of N2:N71 bp. With a large excess of G1**:**C72 bp among tRNA species with different coding specificity, in each species Kingdom, this bp can be inferred to predate species divergence from the Last Common Ancestor (LCA), more than 3.5×10^9 years ago[35,36], and earlier diversification of tRNA coding specificity deep within the pre-LCA era[7]. At the second acceptor stem site, G2**:**C71 and C2**:**G71 accounted for 39 and 44 per cent of bp,

A



B

bp frequency (per cent)

↓N1 \ N72→	A	G	U	C	Total
A	X	X	2.08	0	2.08
G	X	X	2.33	83.69	86.02
U	7.10	0	X	X	7.10
C	0	4.80	X	X	4.80
Total	7.10	4.80	4.41	83.69	100.00

↓N2 \ N71→	A	G	U	C	Total
A	X	X	3.69	0	3.69
G	X	X	5.79	39.09	44.88
U	6.62	0.66	X	X	7.28
C	0	44.17	X	X	44.17
Total	6.62	44.83	9.48	39.09	100.02

↓N3 \ N70→	A	G	U	C	Total
A	X	X	10.57	0	10.57
G	X	X	7.36	31.25	38.61
U	13.60	0.72	X	X	14.32
C	0	36.50	X	X	36.50
Total	13.60	37.22	17.93	31.25	100.00

C

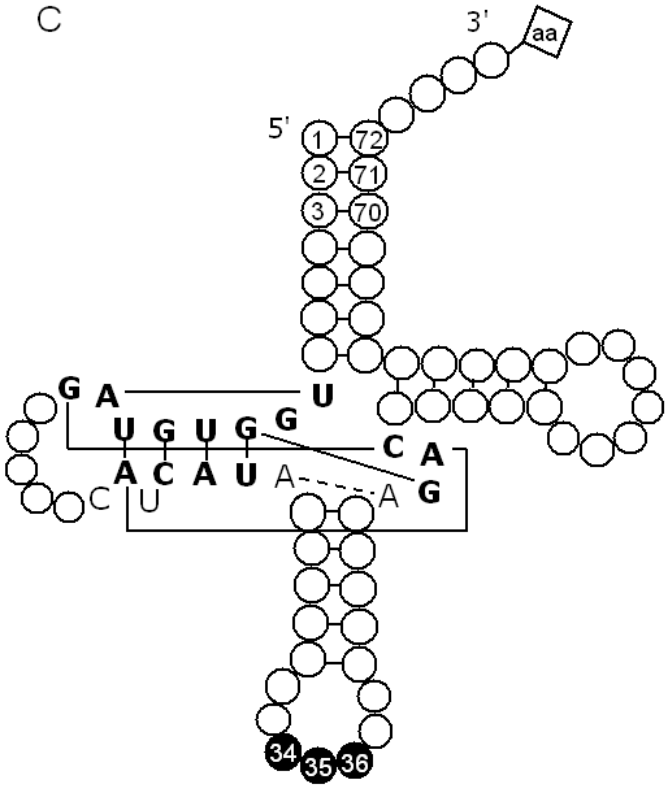


Figure 1. Nucleotide distribution in tRNA acceptor stem terminal triplets and anticodon. (**A**) Acceptor stem terminal triplets, N1-N3 and N72-N70, show anti-parallel complementarity at these G,C rich sites (P = 0.991, goodness of fit). The 5'-terminal triplet and anticodon exhibit no identity, or complementarity. Prokaryote tRNA N34-site is A-deficient (black bar) unlike eukaryote tRNA (white bar). Mean frequencies were based on nucleotide mole ratios in ref. 30. (**B**) Base-pair distribution evaluated for acceptor stem terminal triplet. G2:C71 and C2:G71 bp are prevalent at the second acceptor stem site, producing N2-base complementarity when counter-aligned in tRNA pairs. (**C**) tRNA cloverleaf diagram showing sites examined. Letters and lines refer to bases and their interactions in adaptor core; illustrated by a type I-D tRNA[7,21].

respectively (Fig. 1A,B). As each N2:N71 bp orientation is broadly equifrequent at this GC rich site, related tRNA species could be expected to share co-aligned N2:N71 bp. Conversely, counter-aligned bp appear likely in unrelated tRNA. N3:N70 bp also have similar G**:**C and C**:**G mole ratios (Fig. 1A,B), but they represent less than 68 per cent of bp. Higher A,U levels account for this (Fig. 1). (GC) bp at N3:N70 should consequently correlate more weakly with tRNA identity. GC levels were found to be only 2.7 per cent higher in pre-divergence tRNA[7] than in extant adaptors[30], allowing comparisons between base profiles in tRNA from both eras.

Code domains of codon-cognate and anticodon-cognate tRNA

Among 24 pairs of codon-cognate and anticodon-cognate tRNA in Fig. 2, only 3 pairs shared the same code domain: Ile^{7}-**D1**/Asn^{2}-**D1** (3'-UAA:UUA-5'); Ser^{4}-**QD2**/ Gly^{5}-**QD2** (3'-AGG:CCU-5'); Thr^{6}-**D1**/ Arg^{9}-**D1** (3'-UGC:GCA-5'), where a bold letter and number specifies the code domain of each tRNA. Among 19 pairs of codon-cognate/anticodon-cognate tRNA with specifiable N2-bases, furthermore, only 4 pairs displayed N2-base identity: Phe^{11}-**C2**/Glu^{1}-**C2** (3'-AAG:CUU-5'); Leu^{7}-**C2**/Glu^{1}-**C2** (3'-GAG:CUC-5'); Pro^{4}-**G2**/Trp^{14}-**G2** (3'-GGU:ACC-5'); and, Ala^{2}-**G2**/Cys^{5}-**G2** (3'-CGU:ACG-5'), where a bold letter and number denote the N2-base that is the consensus N2-base of the common ancestor for each tRNA species in Bacteria, Archaea, or Eukarya; based on ref. 11.Different core groups[7,21] in 17/24 tRNA pairs provide further evidence of heterogeneity between

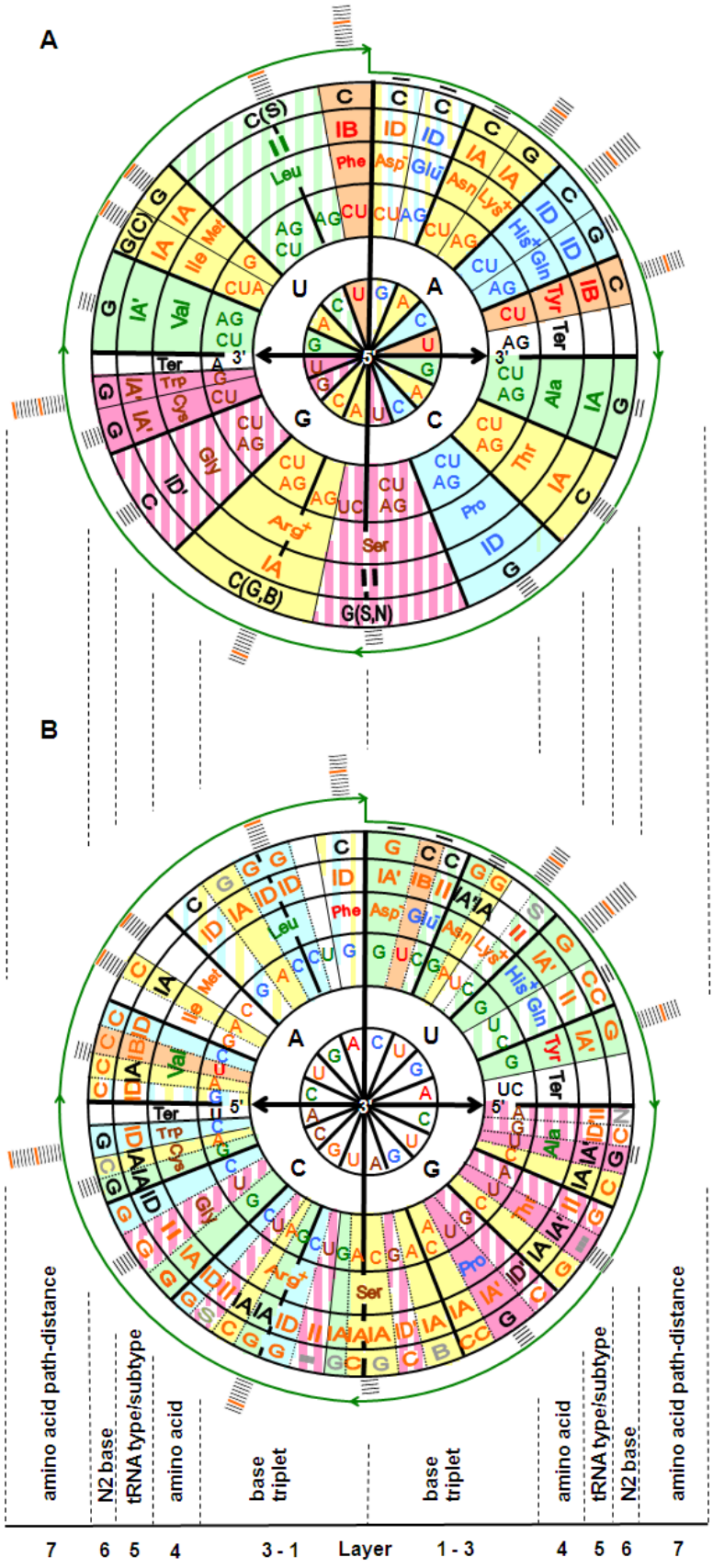

A
B
amino acid path-distance
N2 base
tRNA type/subtype
amino acid
base triplet
base triplet
amino acid
tRNA type/subtype
N2 base
amino acid path-distance
7
6
5
4
3 - 1
Layer
1 - 3
4
5
6
7

Figure 2. Code domain distribution of codon-cognate and anticodon-cognate tRNA. (**A**) Standard code map[19] showing codons (5' to 3' base, layers 1-3), amino acids (layer 4), codon-cognate tRNA core structure group (layer 5) and N2-base (layer 6), and amino acid path-distance stacks (number of reaction steps in synthesis path, layer 7). Inside green arrows are stacks for amino acids with 1-7 step paths (1 bar per step, orange bar at a 7-step span). Outer stacks are for amino acids with 9-14 step paths. Arrows show direction of code expansion, by recruiting unassigned triplets, inferred from increases in earlycomer amino acid path-distance (1-7 steps). Long-path (9-14 step) amino acids generally have only 1-2 triplets, in a set-of-four triplets shared with a same-family, short-path amino acid, or chain termination signal, consistent with codon capture by overprinting in the post-expansion code. The code is subdivided[7] into five domains (D) and three quasidomains (QD): D1 contains OAA-derived amino acids, type-IA tRNA, and ANN codons (orange letters, yellow background). D2, KG-derived amino acids, type-ID tRNA, and CNN codons (blue letters/turquoise background). D3, Pyr-derived amino acids, type-IA and -IA' tRNA, and GNN codons (green letters/light-green background). D4, PG-derived amino acids, type-IA' tRNA, UGN codons (brown letters/rose background). D5, PEP and EP-derived amino acids, type-IB tRNA, UNN codons (red letters/tan background). QD1, OAA- or KG-derived diacid amino acids, type-ID tRNA, GAN codons (orange and blue letters/yellow and azure striped background). QD2, PG-derived amino acids, type-ID' and –II tRNA, UCN, RGN codons (brown letters/rose striped background). And, QD3, Pyr-derived amino acid Leu, type-II tRNA isoacceptors, UUR, CUN codons (green letters/striped light-green background). (**B**) An anticode map showing anticodons (3' to 5'), amino acids, anticodon-cognate tRNA core group and N2-base, and amino acid path-distance stacks. Reduced tRNA A34 frequency lowered the number of coding triplets in this map. A conspicuously fragmented (non-periodic) background pattern illustrates that tRNA with complementary coding triplets, as exemplified by anticodon-cognate tRNA versus codon-cognate tRNA species (upper code map), are generally from different code domains. An orange letter in layer 5 or 6 indicates an anticodon-cognate tRNA possessed a different core group or N2-base, respectively, from its corresponding codon-cognate tRNA; a black letter designates identity, and a grey letter or dash (layer 6) an ambiguous match.

codon-cognate and anticodon-cognate tRNA (Fig. 2, layer 5). Seven of 24 tRNA pairs had identical core groups: 6 type-**IA** tRNA pairs, Ile^{7}/Asn^{2} (3'-UAA:UUA-5'); Val^{4}/Asn^{2} (3'-CAA:UUG-5'); Thr^{6}/Cys^{5}

(3'-UGU:ACA-5'); Thr^{6}/Arg^{9} (3'-UGC:GCA-5'); Ala^{2}/Cys^{5} (3'-CGU:ACG-5'); and, Ala^{2}/Arg^{9} (3'-CGC:GCG-5') and 1 type-**ID** pair, Pro^{4}/Gly^{5} (3'-GGG:CCC-5'), where a bold roman numeral and letter identify the core group. Combining probabilities for domain, N2-base, and core group mismatches between codon-cognate and anticodon-cognate tRNA reveals they form extremely heterogeneous pairs versus same-domain pairs, which matched tRNA within Fig. 2A ($P = 1.81 \times 10^{-15}$). Exclusion of mixed-domain pairs from the latter set accounts for this result. Inspection of the code map (Fig. 2A) reveals only 6/24 same-domain tRNA pairs differ at N2 (2 Lys^{10}-**G2**:Asn^{2}-**C2**; 2 His^{13}-**C2**:Gln^{2}-**G2**; 2 Gly^{5}-**C2**:Ser^{4}-**G2**) and 2 pairs had different core groups (2 Gly^{5}-**ID'**:Ser^{4}-**II**).

Regrouped tRNA pairs

Reversing the formation of mixed-domain pairs among tRNA with complementary anticodons (Fig. 2), by regrouping them into same-domain pairs, will be shown here to extinguish elevated N2-base complementarity and other expressions of heterogeneity. A set of 24 complementary tRNA pairs, uniquely defined by canonical H-bond interactions between anticodon bases [based on ref. 11], contained 21 mixed-domain and 3 same-domain (including quasidomain) pairs (Table 1). Among 19 complementary-set pairs with a specifiable N2-base, only 4 pairs shared the same N2-base. The remaining 15 tRNA pairs had complementary N2-bases (counter-aligned N2:N71 bp). In addition, only 6 complementary-set pairs contained nearest-neighbor anticodons. Anticodons in the remaining 18 pairs were non-contiguous. Mean (± s.e.m.) sequence identity[7] at non-universal tRNA sites, jointly conserved across species Kingdoms, was only 3.65 ± 0.45 quarts per pair in this set. Regrouping complementary-set tRNA produced 22 same-domain pairs;.mixed-domain pairs were excluded, by the pairing rule, from this set. However, 15/19 pairs displayed N2-base identity, and 17/22 pairs had contiguous anticodons. Mean quaternary sequence identity among same-domain tRNA pairs

	Complementary anticodons						Same code domain					
	Amino acids	Anticodons (3' - 5')	N2-base	aaRS class	Identity (quarts)		Code domain	Amino acids	Anticodons (3' - 5')	N2-base	aaRS class	Identity (quarts)
1	Phe^{11}:Glu^{1}	AAG:CUU	**C:C**	II:I	0.6		D1	**Asn^{2}:Thr^{6}**	**UUG:UGG**	**C:C**	**II:II**	7.5
2	Leu^{7}:Gln^{2}	AAC:GUU	C:G	**I:I**	4.0			**Asn^{2}:Lys^{10}**	**UUG:UUU**	C:G	**II:II**	16.0
3	Leu^{7}:Lys^{10}	GAA:UUC	(S:G)	I:II	3.9			**Thr^{6}:Arg^{9}**	**UGU:UCU**	(C:B)	II:I	10.0
4	Leu^{7}:Glu^{1}	GAG:CUC	**C:C**	**I:I**	2.3			**Thr^{6}:Arg^{9}**	**UGC:UCC**	**C:C**	II:I	6.0
5	Leu^{7}:Gln^{2}	**GAC:GUC**	C:G	**I:I**	0.2			**Ile^{7}:Arg^{9}**	**UAU:UCU**	(C:B)	**I:I**	9.0
6	**Ile^{7}:Asn^{2}**	**UAA:UUA**	*G:*C	I:II	3.9			**Arg^{9}:Arg^{9}**	**UCC:GCC**	**C:C**	**I:I**	7.0
7	Val^{4}:Asn^{2}	CAA:UUG	*G:C	I:II	3.4		D2	**Gln^{2}:Pro^{4}**	**GUC:GGC**	**G:G**	I:II	3.0
8	Val^{4}:Asp^{1}	**CAG:CUG**	G:C	I:II	0.9			**Gln^{2}:Pro^{4}**	**GUU:GGU**	**G:G**	I:II	4.0
9	Val^{4}:Tyr^{11}	CAU:AUG	G:C	**I:I**	4.0			**Pro^{4}:His^{13}**	**GGG:GUG**	G:C	**II:II**	2.2
10	Val^{4}:His^{13}	CAC:GUG	G:C	I:II	2.0		D3	**Ala^{2}:Val^{4}**	**CGA:CAA**	***G:*G**	II:I	9.4
11	Ser^{4}:Arg^{9}	AGA:UCU	(-:B)	II:I	5.1			**Ala^{2}:Val^{4}**	**CGG:CAG**	**G:G**	II:I	9.4
12	**Ser^{4}:Gly^{5}**	AGG:CCU	G:C	**II:II**	9.9			**Ala^{2}:Val^{4}**	**CGU:CAU**	**G:G**	II:I	5.0
13	Ser^{4}:Arg^{9}	AGC:GCU	(S:G)	II:I	5.0			**Ala^{2}:Val^{4}**	**CGC:CAC**	**G:G**	II:I	3.0
14	Pro^{4}:Arg^{9}	GGA:UCC	*G:C	II:I	1.0		D4	**Cys^{5}:Trp^{14}**	**ACG:ACC**	**G:G**	**I:I**	4.0
15	Pro^{4}:Gly^{5}	GGG:CCC	G:C	**II:II**	2.7		D5	**Phe^{11}:Tyr^{11}**	**AAG:AUG**	**C:C**	II:I	10.0
16	Pro^{4}:Trp^{14}	GGU:ACC	**G:G**	II:I	3.0		PD1	**Asp^{1}:Glu^{1}**	**CUG:CUU**	**C:C**	II:I	7.0
17	Pro^{4}:Arg^{9}	**GGC:GCC**	G:C	II:I	3.2		PD2	**Ser^{4}:Gly^{5}**	UCA:CCU	G:C	**II:II**	8.0
18	Thr^{6}:Ser^{4}	**UGA:UCA**	C:G	**II:II**	4.9			**Ser^{4}:Gly^{5}**	UCG:CCC	(N:C)	**II:II**	6.0
19	Thr^{6}:Cys^{5}	UGU:ACA	(C:-)	II:I	7.0			**Ser^{4}:Gly^{5}**	AGG:CCG	G:C	**II:II**	10.4
20	**Thr^{6}:Arg^{9}**	UGC:GCA	C:G	II:I	6.6			**Ser^{4}:Ser^{4}**	AGG:UCG	**G:G**	**II:II**	1.0
21	Ala^{2}:Ser^{4}	CGA:UCG	(*G:N)	**II:II**	2.3		PD3	**Leu^{7}:Leu^{7}**	AAU:GAG	**C:C**	**I:I**	8.3
22	Ala^{2}:Gly^{5}	**CGG:CCG**	G:C	**II:II**	2.7			**Leu^{7}:Leu^{7}**	**AAC:GAC**	**C:C**	**I:I**	9.4
23	Ala^{2}:Cys^{5}	CGU:ACG	**G:G**	II:I	4.8		-	Glu^{1},Pro^{4}				
24	Ala^{2}:Arg^{9}	CGC:GCG	G:C	II:I	4.0		-	Cys^{5},Arg^{9}				
	3	**6**	**4**	**9**	3.65			**22**	**17**	**15**	**12**	7.07
	21	18	15	15	± 0.45			0	5	4	10	± 0.74

Table 1. tRNA sets with complementary anticodons versus regrouped same-domain pairs. D1-D5 and QD1-QD3 designate tRNA code domains and quasidomains. Same-domain pairs in each set are in bold letters. Superscripts indicate amino acid synthesis path-distances. Contiguous anti-codons (single base difference) are in bold letters. N2-base identity is marked by bold letters; other pairs have complementary N2-bases. S denotes C2 or G2; B is C, G or U; N, any base; -, tRNA from a single species Kingdom, and, *, single-Kingdom tRNA with an A34 and same N2-base as G-bearing isoacceptors. Parenthesis enclose an ambiguous N2-base. Bold letters show paired tRNA with same aaRS class. Base identity for conserved trace of pre-divergence tRNA sequences is given in quaternary units (quarts). Under each column is the total, or mean (± s.e.m.). Complementary set tRNA pairs have fewer same-domain pairs. ($P = 2.91 \times 10^{-10}$), contiguous anticodons ($P = 4.70 \times 10^{-4}$), identcal N2-bases ($P = 4.52 \times 10^{-4}$), and less pre-divergence sequence identity ($P = 1.04 \times 10^{-4}$).

was 7.07 ± 0.74 quarts per pair. tRNA in same-domain pairs were significantly more homologous than in complementary set pairs ($P = 1.57x10^{-9}$, combined probability). This result cannot be attributed to the choice of tRNA species, as complementary-set tRNA were used to form the same-domain pairs. Differences observed in sequence identity, anticodon contiguity, N2-base correspondence, and domain homology directly relate to tRNA structure and code evolution. The distribution of class I and II aaRS, on the other hand, showed no significant difference between complementary and same-domain pairs (Table 1). A distinction thus arises between the products of translation (aaRS) and a mediator of protein synthesis (tRNA) in their relation to code structure.

Correlates of tRNA identity

The structural and phylogenetic (post-divergence variations filtered-out) homology displayed by same-domain tRNA[7] implies specific regions of the code are occupied by tRNA species that diversified from a common ancestral tRNA. Alternative pairing criteria, such as matching tRNA with complementary anticodons[11], could be anticipated to produce arbitrary tRNA pairs with respect to code structure. According to this, randomly regrouping tRNA in the same-domain set (Table A2) should show elevate N2-base complementary and other expressions of heterogeneity (Table 1). Nearly identical mixed- and same-domain pair frequencies among complementary and random-domain tRNA sets (Fig. 3A) support this. Both distributions contain a large excess of mixed-domain pairs, which contrasts with their exclusion from the same-domain set. About equal numbers of pairs with same- and complementary-N2-bases occurred in the random-domain set, placing it midway between complementary and same-domain sets (Fig. 3B). Differences with both these sets showed borderline significance ($P = 0.049$, combined probability). Randomness produced by an arbitrary tRNA pairing rule, with respect to code structure, clearly contributed to complementary-set N2-base

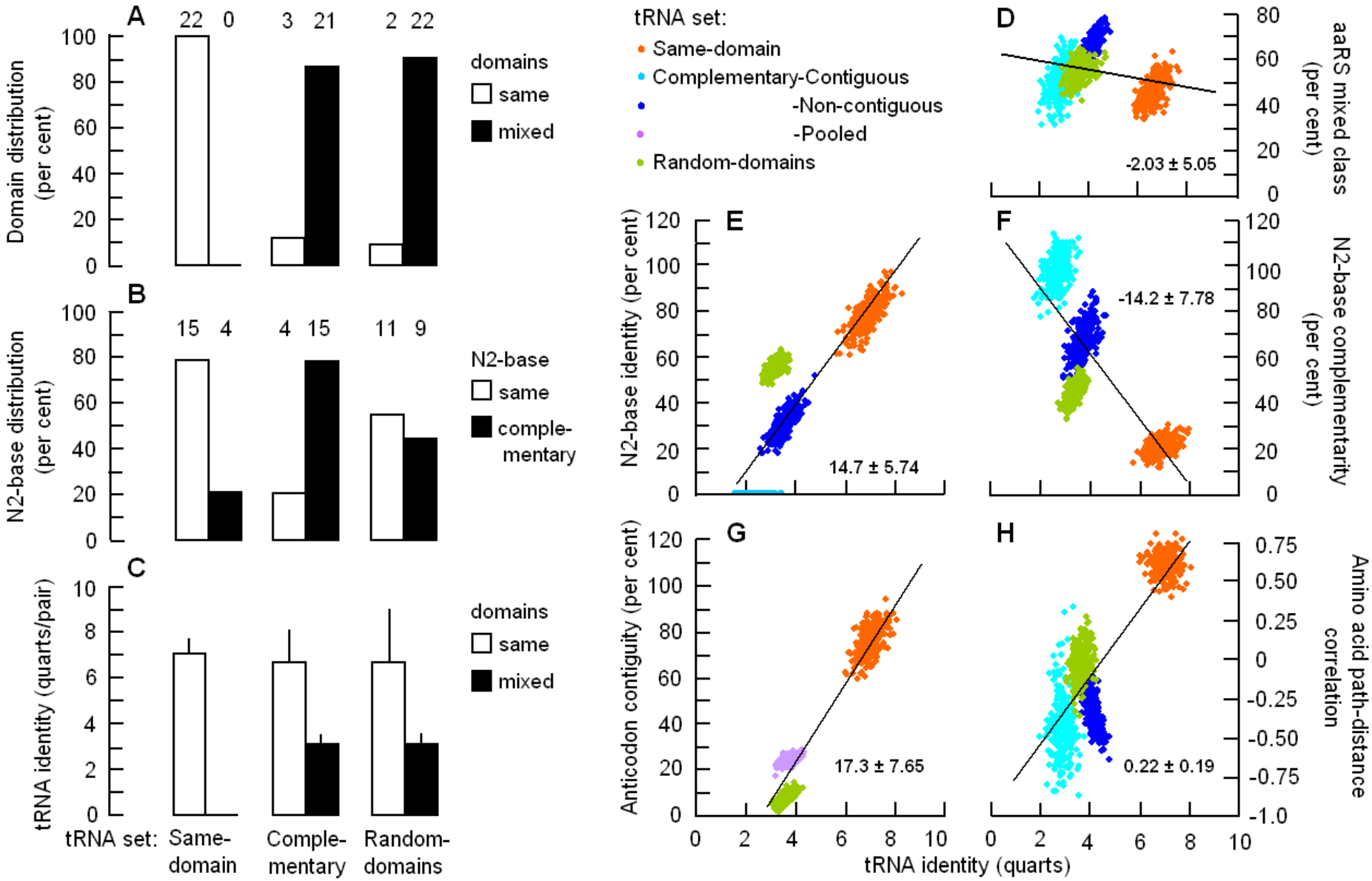

A
22 0
3 21
2 22
Domain distribution (per cent)
domains
same
mixed
B
15 4
4 15
11 9
N2-base distribution (per cent)
N2-base
same
comple-mentary
C
tRNA identity (quarts/pair)
domains
same
mixed
tRNA set: Same-domain
Comple-mentary
Random-domains
tRNA set:
Same-domain
Complementary-Contiguous
-Non-contiguous
-Pooled
Random-domains
D
-2.03 ± 5.05
aaRS mixed class (per cent)
E
14.7 ± 5.74
N2-base identity (per cent)
F
-14.2 ± 7.78
N2-base complementarity (per cent)
G
17.3 ± 7.65
Anticodon contiguity (per cent)
H
0.22 ± 0.19
Amino acid path-distance correlation
tRNA identity (quarts)

Figure 3. Correlates of pre-divergence sequence identity in same-domain, complementary, and random-domain tRNA pairs. (A) Complementary and random-domain sets mainly pair tRNA from different code domains. Same-domain pairs exclude mixed-domain tRNA and the resulting difference is significant ($P = 6.47 \times 10^{-13}$, Fisher's test). Each set constituted a different arrangement of the same tRNA. Above each column is the number of tRNA pairs per set. (B) 15/19 complementary set pairs had complementary N2-bases versus 4/19 among same-domain pairs ($P = 4.52 \times 10^{-4}$) and 9/20 in random pairs ($P = 6.44 \times 10^{-2}$, χ^2 test). (C) Same-domain pairs in each of the three tRNA sets show comparable mean (± s.e.m.) pre-divergence tRNA sequence identity, specified in quaternary units (quarts). Mixed-domain pairs, in the complementary and random set, show significantly lower tRNA sequence identity ($P < 1 \times 10^{-2}$, ANOVA, Tukey's multiple comparison test). (D) Pairs having mixed class aaRS occur with comparable frequency in four tRNA sets (same-domain, contiguous and non-contiguous complementary anticodons, and random-domains) with indicated pre-divergence tRNA sequence identity. Mean linear regression slope is given, with its 98 per cent (non-parametric) confidence interval, which spans zero. (E and F) N2-base identity and complementarity show a positive and negative dependence, respectively, on pre-divergence tRNA sequence identity across the four sets examined. (G) tRNA pairs having contiguous anticodons increased with sequence identity over same-domain, complementary, and random sets. (H) Pearson correlation between cognate amino acid path-distances increased with tRNA sequence identity.

complementarity, as anticipated. Complementary set N2-base complementarity exceeded that in the random-set, however. A second source of N2-base complementarity was therefore sought (see below). Sequence identity among same-domain pre-divergence tRNA (Fig. 3C) exceeded that of mixed-domain pairs, in all three tRNA sets ($P = 1.90 \times 10^{-6}$, OWAV and t test). As mixed-domain pairs exceeded same-domain pairs by 9:1 in complementary and random-domain sets (Fig. 3A), both displayed low mean sequence identity. The nucleotide profile at acceptor stem N2:N71 bp (Fig. 1A,B) suggests elevated N2-base complementarity in complementary and random-domain sets accompanies high sequence heterogeneity (Fig. 3B,C).

N2-base complementarity expected from GC mole ratios at this site (Table 1, Table A2) was 56.8 and 51.0 per cent, respectively, for complementary and random tRNA sets. Since 78.9 and 45.0 per cent of pairs in the respective sets had complementary N2-bases, randomness from an arbitrary

tRNA pairing rule could account for only part of complementary-set complementarity. All 6 tRNA pairs, in a subset with contiguous anticodons, possessed complementary N2-bases; two additional pairs with contiguous (anti-parallel) complementary anticodons, 3'-AGU:**ACU**-5' and 3'-AAU:**AUU**-5', included a chain-termination signal (bold letters) and were discounted. Among complementary set tRNA pairs with non-contiguous anticodons, 9/13 pairs had complementary N2-bases. They show acceptable agreement with the expected value (7.5/13) for this subset and for random-domain pairs (10/20), based on GC mole ratios at N2. 100 per cent (6/6) N2-base complementarity among contiguous-anticodon pairs of the complementary set accompanied a pre-divergence tRNA sequence identity of only 2.64 ± 0.73 quarts per pair (Table A2). This subset contained a single same-domain pair, Ile^{7}:Asn^{2} (3'-UAA:UUA-5'). Both tRNA in this pair contain an A34. This suggests they are of post-divergence vintage, given the restricted distribution of A34-bearing tRNA between species Kingdoms (Fig. 1A). The mixed-domain tRNA pairs in this subset combine synthetically unrelated amino acids and contiguous coding triplets, to form a counter-domain combination. Four of them, lacking an A34, had a sequence identity of only 1.75 ± 0.72 quarts/pair. Hence, it appears they were the target of strong selection forces[37] directed at error-minimization[9] in amino acid synthesis and translation, by enhancing molecular recognition based on differences with same-domain tRNA.

The distribution of class I and II aaRS among tRNA pairs in four distinct tRNA sets showed no dependence on sequence identity. Figure 3D shows a linear slope of -2.03 per cent/quart with 98 per cent confidence limits (± 5.05) that span zero. This agrees with results in Table 1 showing the frequency of mixed-class aaRS pairs to be independent of sequence identity among pre-divergence tRNA species. Clusters of pairs with mixed-class aaRS in each set, accordingly, overlap on the vertical axis of Fig. 3D. On the horizontal axis, by contrast, distinct tRNA identity exist. Sequence identities for same-domain pairs project onto the horizontal axis between 6 – 8 quarts. Random-domain and contiguous and non-contiguous complementary set identities, on the other hand, cluster in the vicinity of 3 - 5 quarts. As each cluster contains 250 means, from 1,000 randomly generated samples of the observed pair-identity distribution[32], Fig. 3D shows sequence identity among pre-divergence tRNA

species, in same-domain pairs, exceeds pair identity in random-domain, and contiguous- and non-contiguous-complementary sets ($P \leq 4x10^{-3}$). Figures 3E-H show same-domain tRNA identity exceeded pair identity for other sets, in each correlation examined, and multiple comparison tests confirmed this relation. Figures 3E,F show sequence identity among pre-divergence tRNA is a determinant of both N2-base identity and complementarity. Acceptor stem N2:N71 bp co-aligned at rate of 14.7 ± 5.74 per cent/quart (mean ± 98 per cent confidence interval) in these four tRNA sets. Conversely, bp at this site counter-aligned at -14.3 ± 7.73 per cent/quart identity among these sets. From coefficient of determination values for each regression, the linear dependence of N2-base identity on tRNA sequence identity accounted for 67.9 per cent of total variability. The negative dependence of N2-base complementarity on sequence identity accounted for 60.2 per cent of variability. Nearest-neighbor anticodon frequency (Fig. 3G) and Pearson correlation for cognate amino acid synthesis path-distances (Fig. 3H), among paired tRNA, show a linear dependence on tRNA sequence identity of 17.3 ± 7.65 per cent/quart and 0.22 ± 0.19 per quart, respectively. Their respective dependence on tRNA identity accounted for 94.4 and 73.9 per cent of total variability

DISCUSSION

Genetic code domains of region-specific, time-ordered codon assignments to amino acids, from the same precursor, conveyed by related tRNA[2,4-7,19], have been demonstrated to clarify why tRNA species, with complementary anticodons, exhibit complementarity at a remote acceptor stem site. Two domain-related sources of N2-base complementarity were identified. tRNA in the complementary set preferentially formed mixed-domain pairs having non-neighboring coding triplets (Figs. 2, 3). With phylogenetically and structurally related tRNA species residing within each code domain[7], mixed-domain pairs generally match unrelated tRNA. Reduced sequence identity among pre-species divergence tRNA (Table 1), as anticipated from the distinctive nucleotide distribution at the acceptor stem N2-site, is coupled to high base complementarity (counter-aligned N2:N71 bp) at this GC rich site (Figs. 1, 3F). In principle, any arbitrary pairing rule with respect to code structure would produce a

tRNA set having comparable domain, sequence, and N2-base heterogeneity. This was illustrated with a set of randomly matched tRNA, formed with regrouped complementary set tRNA species (Table A2). Conversely, regrouping complementary set tRNA into same-domain pairs increased pre-divergence tRNA sequence identity, and this accompanied increased base identity and reduced complementarity at N2 (Table 1, Fig. 3E,F). In addition, error-minimizing selection forces[9,37] evoked by counter-domain tRNA pairs, cognate with nearest-neighbor coding triplets but charged by amino acids from different synthesis families, are considered to have significantly elevated N2-base complementarity in the complementarity set. All 6 tRNA pairs in this subset exhibited complementarity at N2. Moreover, they had the lowest observed sequence identity, 2.64 ± 0.74 quarts/pair (mean ± s.e.m.). The magnitude of the difference between mixed- versus same-domain tRNA, with contiguous anticodon, is apparent on noting that the latter displayed sequence identity of 7.17± 0.85 quarts/pair (from Table 1) and N2-base complementarity of only 13.3 per cent (2/15 pairs with specifiable N2-bases).

By adding to the number of code features attributable to its domain structure, marked by distinct sibling amino acids/contiguous codons/related tRNA combinations (Table A1), the correlation between complementary anticodons and complementarity at the remote acceptor stem N2 site[11] provides additional evidence for occurrence of an extensive network of tRNA-dependent amino acid synthesis pathways throughout code formation[7]. The central role of tRNA in forming the code is further supported by the dependence of anticodon contiguity and cognate amino acid path-distance correlation on pre-divergence tRNA sequence identity (Fig. 3G,H). As 5 code domains and 3 quasidomains incorporate all amino acids in the standard set, their tRNA, and codons (Fig. 2A), amino acid synthesis can be inferred to have extensively utilized tRNA cofactors during code formation. Code structure accordingly excludes direct amino acid recognition by synthetases (ribozymal, enzymic) from playing a significant role in shaping the genetic code[7]. The scope of the reliance on tRNA cofactors conforms with existence of a tRNA-based mechanism for specifically matching amino acids and their codons, before aaRS appeared. Coding fidelity during translation, in

either era, rests on identity elements within tRNA that enable recognition of its amino acid specificity. Rather than recruit a synthetase to catalyze attachment of an activated amino acid, a pre-aaRS tRNA cofactor evidently recruited specific catalysts in the synthesis of its cognate amino acid. Asp-tRNA^{Asn} conversion to Asn-tRNA^{Asn}, and Glu-tRNA^{Gln} to Gln-tRNA^{Gln}, in many prokaryotes[24,25], exemplifies tRNA-mediated path-selection in amino acid synthesis. Path-identity elements in the acceptor-, D-, and T-arms of charged prokaryote tRNA^{Asn} and tRNA^{Gln}, absent from tRNA^{Asp} and tRNA^{Glu}, recruit an amide-transferase[38] to catalyze the respective amidation reaction[22,23]. An amide amino acid tRNA cofactor is thus encoded to direct the synthyesis of its cognate amino acid. Likewise, tRNA cofactors in N-formyl-methionine, selenocysteine, and pyrrolysine pathways recruit specific catalysts required for synthesis of their cognate amino acid[39,40]. Takeover of the whole Val^{4} pathway, to complete Ile^{7} synthesis (Fig. A1), suggests an early tRNA recruited a cassette of ribozymes[41] to catalyze this multistep segment of the Ile^{7} pathway. Code structure and tRNA-dependent amino acid synthesis have broadly served here to bring into focus the origin of proteins within RNA-based life forms on the early Earth.

Acknowledgements

The assistance of Lynn M. Hill and Pablo Rubio, University of California San Diego, in formatting the illustrations is gratefully acknowledged.

Appendix

Code Feature	Interpretation
1. Triplets for aa from the same synthesis family tend to cluster in the same code region[42].	Pre-divergence phylogenetics show tRNA species for same-family aa diversified from a common ancestor[7]. Early aa synthesis pathways were concluded to rely on cofactor/ adaptor tRNA with nearest-neighbor anticodons, resulting in contiguous codons being assigned to same-family aa.
2. Probability experiments reveal codon sets-of-four preceded doublets in code evolution[43].	The code conserves evidence for subdivision of codon sets-of-four (3'-base degenerate)[2]**.** Earlycomer aa (2-7 step paths) have 7 of 8 four-sets. In contrast, all 6 latecomer aa (9-14 step paths) share a doublet, or single, with a short-path aa, or stop signal - Arg^{9} also has four-set, CGN.
3. Eight codon sets-of-four encode a single aa and all have a G/C as 5'- and/or mid-base[44].	Codon four-sets lacking a 5'-, or mid-, G/C can be misread at their mid-base, when a Y**:**Y pair forms at the codon-anticodon wobble site[16]. In the PDM, early tRNA species had a U34 (universal bp-forming anticodon 5'-base) from their common ancestor, $tRNA^{(Asp/Glu/Asn/Gln)}_{UUU}$, cognate with a pre-code poly(A) template[2,5]. Modifying U34 can suppress mis-reading by cutting tRNA reading range to a 3'-Y, or 3'-R, doublet. One doublet then becomes a target for subsequent capture.
4. Genetic code antiquity implies that aa formed by reductive organo-synthesis[43,45,46].	aa C-atom oxidation no. decreases linearly with path-distance among 14 code earlycomers (1-7 step paths)[2]. Six latecomer aa (9-14 step paths) show no decrease, as a reliance on reductive aa synthesis evidently ceased with appearance of functional cells at a 7-9 step code age[6].

Table A1. Path-distance model (PDM) interpretation of structural features attributed to the genetic code. Amino acid (aa) pat-distances (superscripts) are specified in Fig. A1.

Code Feature	Interpretation
5. NAN triplets code for hydrophilic aa and NUN triplets code hydrophobic aa[8]. .	NAN triplets encode short-path (1-2 steps) acidic and polar NH_4^+ fixer/N-donor aa, Asp^-, Glu^-, Asn, Gln. On path-distance evidence they formed the first code[2,5]. As aa paths grew more hydrophobic residues were encoded. They would extend the residence-time of early multi-anionic proteins on a cationic mineral surface (Fajans-Paneth rule[45]). Leu[7], Ile[7], Met[7], Val[4] acquired NUN triplets. With mean path lengths of 6.25 steps, the PDM places entry of these hydrophobic aa into the code in late expansion phase. Basic and aromatic aa, Lys^+, Phe, Tyr, $His^{+/0}$, form in 10-13 steps and their distribution supports capture of NAN and NUN doublets by overprinting subdivided four-sets[16], consistent with optimizing aa physical homology[9].
6. During code formation, aa synthesis possibly used tRNA cofactors[43,47].	tRNA-dependent aa synthesis is credited with partitioning the code into domains of contiguous codons read by related tRNA (same core group[7,21]) specific for sibling aa[7,19].
7. Codon mid-base has most, and 3'-base least, coding capacity[48].	Codon mid-base substitutions primarily added 10 aa (2-7 step paths) in code expansion. 5'-bases designated only 4 aa (1-2 step paths) in the earlier NH_4^+ Fixers Code. Late retruitment of the 3'-base added just 6 basic or aromatic aa (9-14 step paths) by overprinting the code[2,5].
8. Codons for same-family aa share a 5'-base[10].	Asn[2] and Gln[2] acquired AAN and CAN triplets, respectively, in the small NH_4^+ Fixers Code. Asn- and Gln-family tRNA retained specificity for ANN and CNN codons during code expansion, by a series of mid-base substitutions (NAN –› NCN –› NGN –› NUN) and extension of these tRNA-dependent aa synthesis pathways[2,5]. tRNA for Ala[2] and sibling Val[4], likewise, acquired specificity for GNN codons, while Ser[4], Cys[5] and Trp[14], together with Phe[11] and Tyr[11] acquired UNN codons.
9. Each of the six smallest aa in proteins acquired a set-of-four codons[10].	Ala[2], Gly[5], Pro[4], Ser[5], Thr[6] and Val[4] have mean mol. wt. of 103 and path-distance of 4.2 steps. Path-distance indicates these aa entered the code early[2,5]. In the PDM, all acquired stable (low error[16]) four-sets, read by U34-bearing tRNA species. This supports a 'first in, best encoded' rule in codon allocation.

Table A1 (continued).

Code Feature	Interpretation
10. tRNA with complementary anticodons have elevated base complementarity at N2[11].	Complementary anticodons are mostly non-contiguous triplets. Hence, they generally form mixed-domain pairs of unrelated tRNA with low pre-divergence sequence identity (Table 1). Base complementarity at the G,C rich N2 site (counter-aligned N2:N71 bp) correlates with sequence heterogeneity (Fig. 3). Randomly paired tRNA also show higher identity-linked N2-base complementarity versus same-domain pairs.
11. Nucleotide-like aa form on long-paths[49].	Purine-like aa His and Trp, with 13 and 14 step paths, were late additions to the code[5]. This places the protein takeover of ribozyme reaction mechanisms late in code formation. tRNA-dependent aa synthesis, another transitional process, likewise persisted into advanced stages of code formation[2].
12. Code distribution of sub-divided four-sets is linked to codon mid-base mis-reading, induced by wobble-site pyrimidine pair[16].	A wobble-site Y**:**Y pair induced mid-base misreading in 8 sub-divided four-sets[16]**:** GAN, AAN, AGN, AUN, CAN, UAN, UGN, UUN. In the PDM, tRNA obtained a U34 (universal bp-forming base) from a pre-code tRNA, cognate with AAA in poly(A)[2]. U34 modification blocks mid-base misreading, but restricts tRNA range from 4 to 2 codons, subdividing the four-set.
13. Frequency of code early-comer aa increases with phylogenetic depth of a residue sequence[50].	Gainer-aa (Ala^{2}, Val^{4}, Gly^{5}, Ile^{7}, Asp^{1}, Ser^{4}, Asn^{2}) form on 1-7 step paths, making them code earlycomers; His^{13} is the sole exception. Four (4/6) loser-aa, Lys^{10}, Phe^{11}, Tyr^{11}, Trp^{14}, form in 9-14 steps, making them latecomers. Pre-divergence aa residue profiles thus show good agreement with the PDM[2,6].
14. Codons for charged aa residues bear a mid-base purine, while codons for hydrophobic aa have a mid-base pyrimidine[51].	Asp^{-}, Glu^{-} acquired GAN, when NAN triplets were assigned in the first code. Lys^{+} and $His^{+/0}$ later overprinted AAN and CAN within their respective code domains. Basic aa, Arg^{9}, likely replaced an Arg-intermediate at CGN and Ser^{4} at AGR[2,52]. Charged aa codons differ by a single base, minimizing mutation risk. Hydrophobic aa Leu^{7}, Ile^{7}, Met^{7}, Val^{4} had clustered at NUN in the late expansion phase code. Subsequent capture of doublet UUY by Phe^{11} completed the hydrophobic cluster.

Table A1 (continued).

Code Feature	Interpretation
15. GNN triplets encode aa that predate ANN-encoded aa, based on changes in aa residue frequency with phylogenetic depth[53].	GNN triplets encode 5 aa (Asp1, Glu1, Ala2, Gly5, Val4) with higher relative frequency in pre-divergence proteins than 7 ANN-encoded aa (Asn2, Ser4, Thr6, Ile7, Met7, Arg9, Lys10). In accord with this, GNN aa have a mean path-distance of 2.6 steps versus ANN aa with 6.4 steps. Although NAN encoded NH4+ fixer/N-donor aa (Asp1, Glu1,Asn2, Gln2), with a 1.5 step mean path-distance, predate the GNN aa[2,19].
16. Similar aa have nearest-neighbor codons, because selection forces acted to minimize deleterious effects from random aa substitutions during code evolution[9,54]. .	Among expansion phase aa (paths < 7 steps), all 4 hydro-philic aa (Asp^1, Glu^1, Asn^2, Gln^2) form in 1-2 steps and cluster on NAN codons. While all 4 hydrophobic aa (Val^4, Ile^7, Met^7, Leu^7) have paths with a mean of 6.25 steps and cluster on NUN codons: Psteps+hydropathy = $1.9x10^{-3}$ (ref. 2,5). Phased addition of aa during code expansion toward a hydrophobic attractor can explain these clusters. Of 4 aa with paths > 9 steps, 3 with NAN codons (Lys^{10}, Tyr^{11}, His^{13}) are hydrophilic (mean path, 11.3 steps), but hydrophobic aa, Phe^{11}, has an NUN doublet. Their distribution fits with a homology optimizing force acting on the post-expansion code. This scalar force would, however, allow any of 24 (4 x 3 (5'-base) + 4 x 3 (mid-base)) codon set combinations. The PDM shows why hydrophilic and hydrophobic aa, respectively, cluster on NAN and NUN triplets[2,5].
17. All four NH4+ Fixer/N-donor aa acquired NAN codons[5].	NH_4^+ fixer/N-donor aa (Asp^1, Glu^1, Asn^2, Gln^2) formed the first PDM code. Back-tracking from their NAN codons suggests proteins originated as random oligopeptides of amide and anionic aa residues, assembled on a pre-code poly(A) template, at the point of entry of N atoms into primal pathways, originating in central metabolism and reliant on charge-attractio[n2,5].
18. Codon mid-base correlates with aa path-distance among aa with 1-7 step paths[2,5].	NAN, NCN, NGN and NUN triplets code for sets of short-path aa (Asp^1, Glu^1,Asn^2,Gln^2), (Ala^2, Thr^6, Pro^4, Ser^4), (Gly^5,Ser^4, Cys^5) and (Val^4,Ile^7, Met^7, Leu^7), with mean path lengths of 1.5 –› 4.0 –› 4.7 –› 6.25 steps, respectively. Expansion from the NH4+ Fixers Code evidently proceeded by recruiting triplets with successive mid-base substitutions[2,5].

Table A1 (continued).

Code Feature	Interpretation
19. 7/8 four-sets code for short-path (1-7 steps) aa. 6/6 long-path aa (9-14 steps), in contrast, have a codon doublet, or single[5].	Short-path aa, Ala[2], Thr[6], Pro[4], Ser[4], Gly[5], Val[4], Leu[6] acquired codon four-sets GCN, ACN, CCN, UCN, GGN, GUN, CUN, leaving only GCN for a long-path aa (Arg[9]). According to this, codons were assigned to early aa as intact sets-of-four. Long-path (latecomer) aa, Lys[10], His[13], Tyr[11], Arg[9], Phe[11], Trp[14] acquired doublets AAR, CAY, UAY, AGR, UUY, UGG and single UGG. Each shares a four-set with a short-path aa, or stop signal, consistent with overprinting the post-expansion code[2,5].
20. Standard code is virtually universal[1].	As aa pathways lengthen a fall-off occurs in the number of codons assigned per step[2], indicative of a slowing in code evolution before the 'universal' code ultimately froze. Thus, the code conserves evidence of a gradual slowing in the tempo of code evolution as aa pathways and genomes grew, increasing the threat of a lethal change to the code[55,56].
21. Acidic aa residues have short paths, and basic aa residues have long paths[2,5].	Diacid aa, Asp^-, Glu^-, form in 1-step, placing them in the first code. Basic aa, Arg^+, Lys^+ and $His^{+/0}$, have 9-13 step paths, consistent with late entry into the code. Early exclusion of positively charged aa mirrors the ubiquity of multianionic molecules in the ancient pathways of central metabolism. Reliance on multianionic metabolites and proteins putatively ceased[2,5], when protocells with a permeable lipid bilayer[56,57], were displaced by functional cells[6].
22. Pre-divergence proteins preferentially conserve short-path aa[6]. .	The aa residue profile at conserved sites in pre-divergence proteins matches the aa alphabet of an earlier code than that of aa at non-conserved sites[6]. Pre-divergence proteins thereby conserve evidence of aa path-distance being a determinant of the time of aa entry into the code.
23. A reconstructed early protein with a stage-5 'code age' has a short acidic segment linked to a cofactor-binding segment[6,58].	A 23-residue ferredoxin antecedent, Pro-Fd-5, with residues from a stage-5.6 aa alphabet (mid-expansion phase code), has a 7 aa negatively charged N-terminus 'foot' linked to a [4Fe-4S] electron transfer center[6,58]. Pro-Fd-5 thus provides a prototype of a protein-adaptor for binding a cofactor to a cationic mineral surface, within a protocell[57], before functional cells evolved[6].

Table A1 (continued)

Code Feature	Interpretation
24. Two unrelated aa, Ser and Leu, have two sets of codons each and uniquely charge type-II tRNA (Fig 2A).	Elevated sequence identity (5.7 quarts) between pre-divergence tRNA-II$^{Ser\ (3'AGU)}$ and tRNA-II$^{Leu\ (3'AAU)}$, a shared core structure group, and nearest-neighbor codons (UC•, UU•) suggest tRNA for Ser[4] was ancestral to Leu[7] tRNA[4]. Relocating identity elements from the anticodon arm to an enlarged variable loop, in type II tRNA, made possible formation of isoacceptors cognate with multiple codon sets[12]. Ser isoacceptors seemingly read multiple UNN and NGN codon sets at some stage during code expansion[5].
25. Diacid aa, Asp^- and Glu^-, form in 1-step, fix NH_4^+ and donate N atoms, are precursors to half aa in proteins, acylate related tRNA-ID, and share the same four-set, GAN[2,7].	Assigning GA• codons to both diacid aa would minimize the effects of 3'-base substitutions. tRNA identities provide evidence of the origin of these codon assignments[2]. Before tRNA specific for each aa in the NH_4^+ Fixers Code had formed, GAN and CAN were found to code for Asp^-, Glu^- and Asn, Gln, respectively[2,7]. The PDM indicates tRNA-IA$^{Asn(UUU)}$ displaced pre-code adaptor, tRNA-ID$^{Asp,Glu,Asn,Gln(UUU)}$, to acquire AAN. Asp^- and Glu^- fix and distribute N atoms, and have catalytic potential, consistent with diacid-aa attached to a proto-tRNA[59] having a pre-translation role.
26. Codon 3'-base is most degenerate[48,1].	With 64 codons for 20 aa and a stop signal, the standard code contains 43 surplus codons. Redundant 3'-bases occur in 39 codons (91 per cent of surplus) spread over 8 four-sets, 1 triplet, and 13 doublets. 3'-bases were not recruited for coding until 6 latecomer aa (9-14 step) entered the code[2,7]. In contrast to other code topologies, notably the 'chromatic code[60], a 3'-base degenerate code suppresses aa substitutions from mutations at this site. It also requires a minimal number of tRNA species (U34-bearing) and assignment of codons in sets-of-four efficiently reduced the risk of lethal (translation-blocking[61]) mutations to unassigned triplets. Mid-base ambiguity during translation of NAN codons[16] by U34-bearing tRNA[2,7], could likewise minimize the risk of unassigned triplets in the first small (NH_4^+ Fixers) code.

Table A1 (continued).

Code Feature	Interpretation
27. Same-family aa charge related tRNA cognate with contiguous codons[7].	tRNA specific for same synthesis family aa generally share the same core group and read contiguous codons[2]. Pre-species-divergence tRNA sequences, with post-species-divergence sequence variations filtered out, reveal tRNA specific for aa, from a common precursor, diversified from a common ancestral tRNA[7]. Omitting to filter-out post-divergence variations (two-thirds of total extant tRNA variations; ref. 20) obscures (see ref. 62) the pre-divergence kinship existing between tRNA species for same family aa.
28. Code structure conserves the imprint of coordinated tRNA diversification, aa synthesis pathway growth, and codon recruitment[7].	Code domains span contiguous codons assigned to same-family aa, conveyed by phylogenetically related pre-divergence tRNA[7]. This is the strongest evidence yet obtained that the growth of aa synthesis pathways (among most conserved pathways known, ref. 63) and code expansion were coordinated with tRNA diversification. The imprint of tRNA-dependent aa synthesis on code structure effectively excludes racemic mixtures of abiogenic aa[64,65] from forming the first generation of genetically encoded proteins.
29. aa specificity of class I and II aaRS conforms with evolution by radiation from precursor (diacid aa) synthetases[7].	aa motifs and signature-segments in aaRS indicate they formed when the aa alphabet was nearing completion[2,5]. Phylogenetic analysis identified Glu-RS-I and Asp-RS-II as early synthetases[66]. aaRS evolution by radiation from Glu-RS-I and Asp-RS-II is supported by their aa substrate distribution. Class I and II aaRS mean substrate molecular weight of 150 (range, 117-204) and 123 (75-165) parallels Glu^1 (146) and Asp^1 (132) molecular weights. Diacid aa ribozymal synthetases putatively charged ancestral tRNA with these precursor aa and were targets of a takeover by Glu-RS-I and Asp-RS-II. Sequence analysis of tRNA, cognate with aa synthetized from diacid aa, indicate $tRNA\text{-}IA^{Asn}$ and $tRNA\text{-}ID^{Gln}$ diversified to form the tRNA cofactors of an extensive network of tRNA-dependent (synthetase-independent) aa synthesis pathways[7].

Table A1 (continued).

Code Feature	Interpretation
30. Direct codon recognition of aa in the early code is favored by NAN and NUN triplets coding for different kinds of aa[8] and by ligand-binding bases in specific anti-aa aptamer[67]	aa path-distance evidence reveals NAN codons were assigned before NUN codons[5,7,19]. Evidence from this source also reveals different kinds of aa were added at different stages in code formation. The phased addition of aa during code expansion thus accounts for Woese clusters[8]. Also contrary to a proto-code with direct aa recognition by codons, NMR spectra show no significant elevation in Arg codon frequency, within an anti-Arg aptamer binding site[68], based on the expected binomial frequency[69].
31. Intermediates in aa synthesis generally bear a free α-carboxyl group (Fig. A1 and ref. 7).	All aa intermediates bear a free α-carboxyl, except His[13] and Trp[14] (steps 12, 13). As this group is masked by a tRNA cofactor in tRNA-dependent aa synthesis, its occurrence in almost all aa intermediates is seen as a relic from the once extensive network of tRNA-dependent aa synthesis pathways[2,7], responsible for shaping code structure[2,5] and pre-divergence tRNA phylogenetics[7].
32. Addition of the α-amine (α-imino) group generally occurs near the end of an aa synthesis path[2].	Seventeen aa (Asp[1], Asn[2], Ile[7], Lys[10], Glu[1], Gln[2], Pro[4], His[13], Ser[4], Cys[5], Gly[5], Trp[14], Ala[2], Val[4], Leu[7], Phe[11], Tyr[11]) in proteins acquire an α-amine (α-imino) in the next-to-penultimate step of synthesis, or later. With most aa intermediates possessing an α-carboxyl, attached to a tRNA cofactor, late addition of an α-amine group would prevent premature aa intermediate translation[2]. Met[7] and sibling Thr[6] are notable exceptions. As Met, or N-formyl-Met, initiate protein synthesis, commonly at an AUG in a template open reading frame, synthesis of regiospecific polypeptides plausibly began with a Met-intermediate, at an advanced stage of the NH_4^+ Fixers Code. Translation from a 5'-initiation site constitutes a precondition for code expansion from the NH_4^+ Fixers Code, linked to regiospecific synthesis of proteins with a larger aa alphabet[4].

Table A1 (continued).

	Random-Domain Set						Subsets of Complementary Set					
	Amino acids	Anticodons (3' - 5')	N2 base	aaRS class	Identity (quarts)		Amino acids	Anticodons (3' - 5')	N2 base	aaRS class	Identity (quarts)
1	**Asn^{2}:Arg^{9}**	UUG:UCU	(C:B)	II:I	10.0			*Contiguous Anticodons*			
2	Asn^{2}:Glu^{1}	UUA:CUU	***C:C**	II:I	1.3	1	Leu^{7}:Gln^{2}	**GAC:GUC**	C:G	**I:I**	0.2
3	Thr^{6}:Gly^{5}	UGU:CCG	**C:C**	**II:II**	2.3	2	**Ile^{7}:Asn^{2}**	**UAA:UUA**	*G:*C	I:II	3.9
4	Thr^{6}:Pro^{4}	**UGC:GGC**	C:G	**II:II**	1.3	3	Val^{4}:Asp^{1}	**CAG:CUG**	G:C	I:II	0.9
5	**Ile^{7}:Arg^{9}**	UAU:GCC	**C:C**	**I:I**	3.4	4	Pro^{4}:Arg^{9}	**GGC:GCC**	G:C	II:I	3.2
6	Arg^{9}:Tyr^{11}	UCC:AUG	**C:C**	**I:I**	4.0	5	Thr^{6}:Ser^{4}	**UGA:UCA**	C:G	**II:II**	4.9
7	Gln^{2}:Gly^{5}	GUC:CCU	G:C	I:II	3.0	6	Ala^{2}:Gly^{5}	**ĊGG:CCG**	G:C	**II:II**	2.7
8	Gln^{2}:Arg^{9}	GUU:UCU	(G:B)	**I:I**	3.0		**1**	**6**	**0**	**3**	2.64
9	Pro^{4}:Leu^{7}	**GGG:GAG**	G:C	II:I	5.1		5	0	6	3	± 0.73
10	Ala^{2}:Arg^{9}	CGA:UCC	*G:C	II:I	4.8						
11	Ala^{2}:Arg^{9}	ĊGG:GCG	G:C	II:I	2.7			*Non-Contiguous Anticodons*			
12	Ala^{2}:Trp^{14}	ĊGU:ACC	**Ġ:G**	II:I	2.2	1	Ṗhe^{11}:Glu^{1}	AAG:CUU	**C:C**	II:I	0.6
13	Ala^{2}:Lys^{10}	ĊGC:UUU	**Ġ:G**	**II:II**	10.0	2	Leu^{7}:Gln^{2}	AAC:GUU	C:G	**I:I**	4.0
14	Cys^{5}:Gly^{5}	ACG:CCC	G:C	I:II	2.7	3	Leu^{7}:Lys^{10}	GAA:UUC	(S:G)	I:II	4.4
15	Phe^{11}:Leu^{7}	AAG:GAC	**C:C**	II:I	4.2	4	Leu^{7}:Glu^{1}	GAG:CUC	**C:C**	**I:I**	2.3
16	Asp^{1}:Ser^{4}	CUG:UCG	(C:N)	**II:II**	1.8	5	Val^{4}:Asn^{2}	CAA:UUG	*G:C	I:II	3.4
17	*Ser^{4}:Val^{4}	UCA:CAA	**G:*G**	II:I	5.7	6	Val^{4}:Tyr^{11}	CAU:AUG	G:C	**I:I**	4.0
18	Ser^{4}:Val^{4}	UCG:CAC	(N:G)	II:I	1.1	7	Val^{4}:His^{13}	CAC:GUG	G:C	I:II	2.0
19	Ser^{4}:Pro^{4}	AGG:GGA	**G:*G**	**II:II**	4.1	8	Ser^{4}:Arg^{9}	AGA:UCU	(-:B)	II:I	5.1
20	Ser^{4}:His^{13}	AGG:GUG	G:C	**II:II**	3.4	9	**Ser^{4}:Gly^{5}**	ȦGG:CCU	G:C	**II:II**	9.9
21	Leu^{7}:Pro^{4}	AAU:GGU	C:G	I:II	0.7	10	Ser^{4}:Arg^{9}	ȦGC:GCU	(Ṡ:G)	II:I	5.0
22	Leu^{7}:Thr^{6}	ȦAC:UGG	**C:C**	I:II	1.9	11	Ṗro^{4}:Arg^{9}	Ġ GA:UCC	*Ġ:C	II:I	1.0
23	Glu^{1}:Val^{4}	ĊUC:CAU	Ċ:G	**I:I**	2.0	12	Pro^{4}:Gly^{5}	ĠGG:CCC	Ġ:C	**II:II**	2.7
24	Cys^{5}:Val^{4}	ACG:CAG	**G:G**	**I:I**	2.1	13	Pro^{4}:Trp^{14}	GGU:ACC	**G:G**	II:I	3.0
						14	Thr^{6}:Cys^{5}	UGU:ACA	(C:-)	II:I	7.0
						15	**Thr^{6}:Arg^{9}**	UGC:GCA	C:G	II:I	6.6
						16	Ala^{2}:Ser^{4}	CGA:UCG	(*G:N)	**II:II**	2.3
						17	Ala^{2}:Cys^{5}	CGU:ACG	**G:G**	II:I	4.8
						18	Ala^{2}:Arg^{9}	ĊGC:GCG	G:C	II:I	4.0
	2	**2**	**11**	**11**	3.45		**2**	**0**	**4**	**6**	4.01
	22	22	9	13	± 0.49		16	18	9	12	± 0.53

Table A2. Significance of anticodon contiguity. All tRNA pairs with contiguous complementary anti-codons had complementary N2-bases (P(binomial) = 1.56×10^{-2}). Non-contiguous complementary pairs show N2-base complementarity, and other forms of heterogeneity, comparable to random-domain pairs (P(combined) = 0.579).

	Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
1	Asp[1]	CUG	ARCHAEA	Archaeglobus Fulg.	DD0340	1	Pro[4]	GGU	ARCHAEA	Archaeglobus Fulg.	DP0341	1	Val[4]	CAC	EUBACT.	Treponema Pallidum	DV1272	1	Leu[7]	AAC	EUBACT.	Acholeplasma Laid.	DL1230
2				Methanococcus Jan.	DD0650	2				Methanococcus Jan.	DP0650	2				Staphylococ. Aure.	DV1481	2				Treponema Pallidum	DL1272
3				Methanococ. Vani.	DD0660	3				Methanococ. Vani.	DP0660	1		CAU	EUKARYA	Plasmodium Falcip.	DV7500	3				Borrelia Burgdorf.	DL1281
4				Methanotherm. Fer.	DD0680	4				Methanotherm. Fer.	DP0680	2				Trypanosoma Brucei	DV7520	4				Staphylococ. Aure.	DL1482
5				Methanococ Voltae	DD0740	5				Methanococ. Voltae	DP0740	3				Leishmania Tarent.	DV7550	5				Helicobacter Pylo.	DL1511
6				Thermococcus Celer	DD0940	6				Thermococcus Celer	DP0940	4				Dictyostelium Dis.	DV7571	6				Bacillus Subtilis	DL1543
7				Haloferax Volcanii	RD0500	7				Haloferax Volcanii	RP0502	5				Saccharomyces Cer.	DV7632	7				E.Coli	DL1662
1			EUBACT.	Mycoplasma Capric.	DD1140	1		GGG		Archaeglobus Fulg.	DP0342	6				Leptomonas Collos.	DV7710	8				Haemophilus Influ.	DL2003
2				Mycoplasma Gen.	DD1150	2				Methanococcus Jan.	DP0651	1				Trypanosoma Brucei	DV7521	9				Synechocystis Sp.	DL2142
3				Mycoplasma Mycoid.	DD1180	3				Haloferax Volcanii	RP0501	2		CAC		Saccharomyces Cer.	RV7631	10				Mycoplasma Capric.	RL1140
4				Mycoplasma Pneumo.	DD1200	1		GGC		Archaeglobus Fulg.	DP0340							11				Mycoplasma Capric.	RL1141
5				Acholeplasma Laid.	DD1230	2				Haloferax Volcanii	RP0500	1	Thr[6]	UGU	ARCHAEA	Archaeglobus Fulg.	DT0342	12				Rhodospiril.Rub.	RL2020
6				Spiroplasma Melif.	DD1260	1		GGU	EUBACT.	Mycoplasma Capric.	DP1140	2				Methanococcus Jan.	DT0650	13				Anacystis Nidulans	RL2100
7				Treponema Pallidum	DD1270	2				Mycoplasma Gen.	DP1150	3				Methanococ. Vani.	DT0660	14				Bacillus Stearo.	RL2120
8				Borrelia Burgdorf.	DD1280	3				Mycoplasma Mycoid.	DP1180	4				Methanotherm. Fer.	DT0680	1		AAU		Mycoplasma Capric.	DL1140
9				Streptomyces Liv.	DD1350	4				Mycoplasma Pneumo.	DP1200	5				Methanococ. Voltae	DT0740	2				Mycoplasma Gen.	DL1150
10				Staphylococ. Aure.	DD1480	5				Spiroplasma Melif.	DP1260	1		UGG		Archaeglobus Fulg.	DT0340	3				Mycoplasma Pneumo.	DL1200
11				Staphylococ. Aure.	DD1481	6				Borrelia Burgdorf.	DP1280	2				Methanococcus Jan.	DT0651	4				Acholeplasma Laid.	DL1232
12				Lactobac.Bulgaric.	DD1500	7				Staphylococ. Aure.	DP1480	3				Methanococ. Vani.	DT0661	5				Treponema Pallidum	DL1270
13				Helicobacter Pylo.	DD1510	8				Lactobac.Bulgaric.	DP1500	4				Thermococcus Celer	DT0940	6				Borrelia Burgdorf.	DL1283
14				Bacillus Subtilis	DD1540	9				Helicobacter Pylo.	DP1511	5				Halobacterium Cut.	RT0380	7				Streptomyces Coel.	DL1310
15				Bacillus Sp. Ps3	DD1570	10				Bacillus Subtilis	DP1540	6				Haloferax Volcanii	RT0501	8				Staphylococ. Aure.	DL1481
16				E.Coli	DD1660	11				E.Coli	DP1660	1		UGC		Archaeglobus Fulg.	DT0341	9				Helicobacter Pylo.	DL1510
17				Haemophilus Influ.	DD2000	12				Salmonella Typhi.	DP1700	2				Haloferax Volcanii	RT0500	10				Bacillus Subtilis	DL1541
18				Haemophilus Influ.	DD2001	13				Photobact. Phosph.	DP1740	1		UGU	EUBACT.	Mycoplasma Capric.	DT1141	11				Bacillus Subtilis	DL1542
19				Haemophilus Influ.	DD2002	14				Aeromonas Hydroph.	DP1780	2				Mycoplasma Gen.	DT1151	12				E.Coli	DL1664
20				Synechocystis Sp.	DD2140	15				Haemophilus Influ.	DP2000	3				Mycoplasma Mycoid.	DT1180	13				Azoarcus Sp.Bh72	DL1950
21				Thermus Thermophi.	RD1580	16				Streptococcus Mut.	DP2070	4				Mycoplasma Pneumo.	DT1202	14				Haemophilus Influ.	DL2000
1			EUKARYA	Plasmodium Falcip.	DD7500	17				Synechocystis Sp.	DP2142	5				Acholeplasma Laid.	DT1230	15				Haemophilus Influ.	DL2001
2				Candida Albicans	DD7600	18				Salmonella Typhi.	RP1702	6				Treponema Pallidum	DT1272	16				Synechocystis Sp.	DL2143
3				Phytophthora Par.	DD7610	1		GGG		Treponema Pallidum	DP1272	7				Borrelia Burgdorf.	DT1280	17				Synechococcus Sp.	DL2150
4				Saccharomyces Cer.	DD7630	2				Helicobacter Pylo.	DP1510	8				Staphylococ. Aure.	DT1480	1		GAU	EUKARYA	Plasmodium Falcip.	DL7500

Table A3. tRNA sources and access numbers (Leipzig database)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
5				Saccharomyces Cer.	DD7631	3				Bacillus Circulans	DP1560	9				Helicobacter Pylo.	DT1510	2				Saccharomyces Cer.	RL7631
6				Schizosaccha.Pom.	DD7640	4				E.Coli	DP1662	10				Bacillus Subtilis	DT1541	1		GAC		Leishmania Tarent.	DL7550
7				Euglena Gracilis	RD7780	5				Synechocystis Sp.	DP2140	11				Thermus Thermophi.	DT1580	2				Candida Tropicali	DL7750
						6				Salmonella Typhi.	RP1701	12				Stigmatella Aurant	DT1630	3				Candida Lusitaniae	DL7760
1	Glu[1]	CUU	ARCHAEA	Archaeglobus Fulg.	DE0340	1		GGC		Treponema Pallidum	DP1270	13				Azospirillum Lipo.	DT1720	4				Pichia Guilliermon	DL7770
2				Pyrococcus Furios.	DE0400	2				Streptomyces Ambo.	DP1360	14				Campylobac. Jejuni	DT1861	5				Candida Albicans	RL7600
3				Methanococcus Jan.	DE0650	3				Mycobact. Tuberc.	DP1400	15				Haemophilus Influ.	DT2000	1		AAC		Saccharomyces Cer.	DL7630
4				Methanococ.Vani.	DE0660	4				E.Coli	DP1661	16				Synechocystis Sp.	DT2141	2				Torulopsis Utilis	RL7650
5				Methanotherm. Fer.	DE0680	5				Synechocystis Sp.	DP2141	17				Mycoplasma Capric.	RT1141	3				Candida Cylindra.	RL7660
6				Haloferax Volcanii	RE0501	6				Salmonella Typhi.	RP1700	18				Mycoplasma Mycoid.	RT1180	1		AAU		Dictyostelium Dis.	DL7570
1		CUC		Archaeglobus Fulg.	DE0341	1		GGU	EUKARYA	Plasmodium Falcip.	DP7500	19				Bacillus Subtilis	RT1540	2				Saccharomyces Cer.	DL7631
2				Ruminobacter Amylo	DE0700	2				Saccharomyces Cer.	DP7630	20				E.Coli	DT1662	3				Saccharomyces Cer.	DL7632
3				Haloferax Volcanii	RE0500	3				Saccharomyces Cer.	DP7631	21				Thermotoga Marit.	DT0990						
1		CUU	EUBACT.	Mycoplasma Capric.	DE1140	4				Torulopsis Utilis	RP7650	1		UGG		Mycoplasma Gen.	DT1150	1	Arg[9]	UCU	ARCHAEA	Archaeglobus Fulg.	DR0344
2				Mycoplasma Gen.	DE1150							2				Mycoplasma Pneumo.	DT1201	2				Methanococcus Jan.	DR0650
3				Mycoplasma Mycoid.	DE1180	1	Ser[4]	UCG	ARCHAEA	Archaeglobus Fulg.	DS0340	3				Treponema Pallidum	DT1271	3				Methanococ.Vani.	DR0660
4				Mycoplasma Pneumo.	DE1200	2				Halobacterium Mar.	DS0440	4				Borrelia Burgdorf.	DT1281	1		UCC		Archaeglobus Fulg.	DR0340
5				Acholeplasma Laid.	DE1230	3				Methanococcus Jan.	DS0651	5				Helicobacter Pylo.	DT1511	1		GCU		Archaeglobus Fulg.	DR0343
6				Treponema Pallidum	DE1271	4				Methanotherm. Fer.	DS0680	6				Bacillus Subtilis	DT1540	2				Methanococcus Jan.	DR0651
7				Borrelia Burgdorf.	DE1280	5				Haloferax Volcanii	RS0500	7				Thermus Thermophi.	DT1581	3				Haloferax Volcanii	RR0502
8				Plesiomonas Shige.	DE1460	1		AGU		Archaeglobus Fulg.	DS0341	8				Stigmatella Aurant	DT1631	1		GCC		Archaeglobus Fulg.	DR0342
9				Haemophilus Ducre.	DE1490	2				Methanococcus Jan.	DS0652	9				E.Coli	DT1660	2				Haloferax Volcanii	RR0500
10				Lactobac.Bulgaric.	DE1500	3				Methanopyrus Kand.	DS0760	10				E.Coli	DT1661	1		GCG		Archaeglobus Fulg.	DR0341
11				Helicobacter Pylo.	DE1510	1		AGC		Archaeglobus Fulg.	DS0342	11				E.Coli	DT1664	2				Methanococcus Jan.	DR0652
12				Helicobacter Pylo.	DE1511	2				Sulfolobus Solfa.	DS0860	12				Listeria Ivanovii	DT1680	3				Halobacterium Cut.	RR0380
13				Lactococcus Lactis	DE1530	3				Halobacterium Cut.	RS0380	13				Listeria Monocyto.	DT1690	4				Haloferax Volcanii	RR0501
14				Bacillus Subtilis	DE1540	4				Haloferax Volcanii	RS0501	14				Pseudomonas Aer.	DT1821	1		UCU	EUBACT.	Mycoplasma Capric.	DR1141
15				Bacillus Subtilis	DE1541	1		AGG		Archaeglobus Fulg.	DS0343	15				Campylobac. Jejuni	DT1860	2				Mycoplasma Gen.	DR1150
16				Bacillus Sp. Ps3	DE1570	2				Methanococcus Jan.	DS0653	16				Haemophilus Influ.	DT2001	3				Mycoplasma Mycoid.	DR1181
17				E.Coli	DE1660	3				Haloferax Volcanii	RS0502	17				Rhizobiumlegumino.	DT2030	4				Mycoplasma Pneumo.	DR1202
18				Aeromonas Hydroph.	DE1780	1		UCG	EUBACT.	Mycoplasma Capric.	DS1141	18				Synechocystis Sp.	DT2142	5				Acholeplasma Laid.	DR1230
19	Glu[1]	CUU	EUBACT.	Haemophilus Influ.	DE2000	2	Ser[4]	UCG	EUBACT.	Mycoplasma Gen.	DS1150	19	Thr[6]	UGG	EUBACT.	E.Coli	RT1660	6	Arg[9]	UCU	EUBACT.	Treponema Pallidum	DR1274
20				Salmonella Enteri.	DE2040	3				Mycoplasma Pneumo.	DS1200	20				E.Coli	RT1661	7				Borrelia Burgdorf.	DR1282

Table A3. (continued)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
21				Synechocystis Sp.	DE2140	4				Acholeplasma Laid.	DS1230	1		UGA		Mycoplasma Capric.	DT1140	8				Helicobacter Pylo.	DR1512
22				E.Coli	RE1661	5				Treponema Pallidum	DS1272	1		UGC		Mycoplasma Gen.	DT1152	9				E.Coli	DR1661
23				E.Coli	RE1662	6				Borrelia Burgdorf.	DS1283	2				Mycoplasma Pneumo.	DT1200	10				Haemophilus Influ.	DR2001
1		CUC		Treponema Pallidum	DE1270	7				Streptomyces Liv.	DS1350	3				Treponema Pallidum	DT1270	11				Synechocystis Sp.	DR2142
2				Streptomyces Rim.	DE1340	8				Helicobacter Pylo.	DS1512	4				Clostridium Aceto.	DT1450	12				E.Coli	RR1662
3				Streptomyces Liv.	DE1350	9				Bacillus Subtilis	DS1542	5				E.Coli	DT1663	13				E.Coli	RR1663
4				Streptomyces Liv.	DE1351	10				Haemophilus Influ.	DS2003	6				Pseudomonas Aer.	DT1820	1		UCC		Mycoplasma Gen.	DR1152
1		CUU	EUKARYA	Plasmodium Falcip.	DE7500	11				Synechocystis Sp.	DS2142	7				Synechocystis Sp.	DT2140	2				Treponema Pallidum	DR1271
2				Dictyostelium Dis.	DE7570	12				Mycoplasma Capric.	RS1140	1		UGU	EUKARYA	Plasmodium Falcip.	DT7500	3				Agrobacter.Tume.	DR1420
3				Saccharomyces Cer.	DE7630	13				Bacillus Subtilis	RS1541	2				Leishmania Tarent.	DT7550	4				Helicobacter Pylo.	DR1513
4				Saccharomyces Cer.	DE7632	14				E.Coli	RS1661	3				Saccharomyces Cer.	DT7632	5				E.Coli	DR1664
5				Schizosaccha.Pom.	DE7640	15				E.Coli	DS1663	4				Eimeria Tenella	DT7680	6				Prochlococcus Mar.	DR1800
1		CUC		Saccharomyces Cer.	DE7631	1		AGU		Mycoplasma Capric.	DS1140	5				Toxoplasma Gondoii	DT7730	7				Streptomyces Vene.	DR2050
2				Schizosaccha.Pom.	DE7641	2				Mycoplasma Gen.	DS1151	1		UGA		Dictyostelium Dis.	DT7570	8				Synechocystis Sp.	DR2141
						3				Mycoplasma Mycoid.	DS1180	2				Saccharomyces Cer.	DT7630	1		GCU		Mycoplasma Gen.	DR1153
1	Asn[2]	UUG	ARCHAEA	Archaeglobus Fulg.	DN0340	4				Mycoplasma Pneumo.	DS1203	3				Saccharomyces Cer.	RT7631	2				Mycoplasma Pneumo.	DR1201
2				Methanococcus Jan.	DN0650	5				Acholeplasma Laid.	DS1231	1		UGC		Trypanosoma Brucei	DT7520	3				Treponema Pallidum	DR1273
3				Methanococ.Vani.	DN0660	6				Spiroplasma Melif.	DS1260	2				Dictyostelium Dis.	DT7571	4				Borrelia Burgdorf.	DR1281
4				Methanotherm. Fer.	DN0680	7				Treponema Pallidum	DS1273	3				Saccharomyces Cer.	DT7631	5				Helicobacter Pylo.	DR1511
5				Halobacterium Cut.	RN0380	8				Borrelia Burgdorf.	DS1281	4				Schizosaccha.Pom.	DT7640	1		GCC		Treponema Pallidum	DR1272
6				Haloferax Volcanii	RN0500	9				Streptomyces Gris.	DS1300							2				E.Coli	DR1660
7				Methanobac.Therm.	RN0620	10				Staphylococ. Aure.	DS1480	1	Met[7]	UAC	ARCHAEA	Archaeglobus Fulg.	DM0340	3				Salmonella Typhi.	DR1700
1			EUBACT.	Mycoplasma Capric.	DN1140	11				Staphylococ. Aure.	DS1481	2				Archaeglobus Fulg.	DM0341	4				Haemophilus Influ.	DR2000
2				Mycoplasma Gen.	DN1150	12				Helicobacter Pylo.	DS1510	3				Archaeglobus Fulg.	DM0342	5				Synechocystis Sp.	DR2140
3				Mycoplasma Mycoid.	DN1180	13				Lactococcus Lactis	DS1530	4				Methanococcus Jan.	DM0651	6				E.Coli	RR1664
4				Mycoplasma Pneumo.	DN1200	14				Bacillus Subtilis	DS1541	5				Methanococcus Jan.	DM0652	7				Aeromonas Hydroph.	DR1780
5				Acholeplasma Laid.	DN1230	15				E.Coli	DS1661	6				Methanothem. Fer.	DM0680	1		GCG		Mycoplasma Gen.	DR1151
6				Treponema Pallidum	DN1270	16				Haemophilus Influ.	DS2000	7				Thermoplasma Acid.	DM0900	2				Mycoplasma Pneumo.	DR1200
7				Borrelia Burgdorf.	DN1280	17				Haemophilus Influ.	DS2002	8				Thermofil. Pendens	DM0960	3				Treponema Pallidum	DR1270
8				Streptomyces Liv.	DN1350	18				Synechocystis Sp.	DS2140	9				Thermotoga Marit.	DM0990	4				Borrelia Burgdorf.	DR1280
9				Streptomyces Liv.	DN1351	19				Mycoplasma Capric.	RS1141	10				Thermotoga Marit.	DM0991	5				Helicobacter Pylo.	DR1514
10				Klebsiella Aeroge.	DN1410	20				Bacillus Subtilis	RS1540	11				Haloferax Volcanii	RM0500	1		GCA		Mycoplasma Capric.	DR1140
11				Lactobac.Bulgaric.	DN1500	21				E.Coli	RS1664	1			EUBACT.	Mycoplasma Capric.	DM1140	2				Mycoplasma Mycoid.	DR1180

Table A3. (continued)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
12				Lactococcus Lactis	DN1530	1		AGC		Mycoplasma Gen.	DS1153	2				Mycoplasma Gen.	DM1150	3				Spiroplasma Melif.	DR1260
13				Bacillus Subtilis	DN1540	2				Mycoplasma Pneumo.	DS1201	3				Mycoplasma Gen.	DM1151	4				Streptomyces Liv.	DR1350
14				Bacillus Subtilis	DN1541	3				Spiroplasma Citri	DS1250	4				Mycoplasma Mycoid.	DM1180	5				Staphylococ. Aure.	DR1480
15				Bacillus Sp. Ps3	DN1570	4				Treponema Pallidum	DS1271	5				Mycoplasma Pneumo.	DM1200	6				Lactobac.Bulgaric.	DR1500
16				E.Coli	DN1660	5				Lactobac.Delbruec.	DS1520	6				Acholeplasma Laid.	DM1230	7				Bacillus Subtilis	DR1540
17				Listeria Ivanovii	DN1680	6				E.Coli	DS1660	7				Acholeplasma Laid.	DM1231	8				E.Coli	DR1663
18				Listeria Monocyto.	DN1690	7				Synechocystis Sp.	DS2143	8				Spiroplasma Melif.	DM1260	9				Salmonella Typhi.	DR1701
19				Haemophilus Influ.	DN2000	8				Lactobac.Bulgaric.	DS1500	9				Treponema Pallidum	DM1270	10				Haemophilus Influ.	DR2002
20				Haemophilus Influ.	DN2001	1		AGG		Mycoplasma Gen.	DS1152	10				Treponema Pallidum	DM1271	11				Haemophilus Influ.	DR2003
21				Haemophilus Influ.	DN2001	2				Mycoplasma Pneumo.	DS1202	11				Borrelia Burgdorf.	DM1280	12				Haemophilus Influ.	DR2004
22				Synechocystis Sp.	DN2140	3				Treponema Pallidum	DS1270	12				Borrelia Burgdorf.	DM1281	13				Synechocystis Sp.	DR2143
23				Azospirillum Lipo.	RN1720	4				Borrelia Burgdorf.	DS1282	13				Staphylococ. Aure.	DM1480	14				E.Coli	RR1660
24				Azospirillum Lipo.	RN1721	5				Helicobacter Pylo.	DS1511	14				Helicobacter Pylo.	DM1510	15				E.Coli	RR1661
1	Asn[2]	UUG	EUKARYA	Plasmodium Falcip.	DN7500	6	Ser[4]	AGG		Bacillus Subtilis	DS1540	15				Helicobacter Pylo.	DM1511	1		UCU	EUKARYA	Plasmodium Falcip.	DR7501
2				Trypanosoma Brucei	DN7520	7				Bacillus Sp. Ps3	DS1570	16				Bacillus Subtilis	DM1540	2				Trypanosoma Brucei	DR7521
3				Trypanosoma Brucei	DN7521	8				E.Coli	DS1664	17				Bacillus Subtilis	DM1541	3				Dictyostelium Dis.	DR7571
4				Tetrahymena Pyrif.	DN7530	9				Haemophilus Influ.	DS2001	18				E.Coli	DM1660	4				Saccharomyces Cer.	DR7631
5				Dictyostelium Dis.	DN7570	10				Haemophilus Influ.	DS2004	19				Photobac. Leiogna.	DM1750	5				Saccharomyces Cer.	RR7632
6				Saccharomyces Cer.	DN7630	11				Clostridium Perfr.	DS2130	20				Haemophilus Influ.	DM2000	1		UCC		Trypanosoma Brucei	DR7520
7				Schizosaccha.Pom.	DN7640	12				Synechocystis Sp.	DS2141	21				Haemophilus Influ.	DM2001	2				Saccharomyces Cer.	DR7632
8				Yersinia Pseudotu.	DN7740	13				Synechococcus Sp.	DS2150	22				Haemophilus Influ.	DM2002	1		GCU		Leishmania Tarent.	DR7551
						14				Bacillus Subtilis	RS1542	23				Haemophilus Influ.	DM2004	1		GCA		Plasmodium Falcip.	DR7500
1	Gln[2]	GUU	ARCHAEA	Archaeglobus Fulg.	DQ0341	15				E.Coli	RS1662	24				Thermus Thermophi.	RM1580	2				Trypanosoma Brucei	DR7522
2				Methanococcus Jan.	DQ0650	16				E.Coli	RS1663	1			EUKARYA	Plasmodium Falcip.	DM7500	3				Leishmania Tarent.	DR7550
3				Methanococ.Vani.	DQ0660	1		UCG	EUKARYA	Plasmodium Falcip.	DS7500	2				Plasmodium Falcip.	DM7501	4				Dictyostelium Dis.	DR7570
4				Methanopyrus Kand.	DQ0760	2				Dictyostelium Dis.	DS7571	3				Dictyostelium Dis.	DM7570	5				Neurospora Crassa	DR7590
1		GUC		Halobacterium Cut.	RQ0380	3				Saccharomyces Cer.	DS7631	4				Saccharomyces Cer.	DM7630	6				Saccharomyces Cer.	DR7630
2				Haloferax Volcanii	RQ0500	4				Candida Cylindra.	RS7664	5				Saccharomyces Cer.	DM7631	7				Schizosaccha. Pom.	DR7640
1		GUU	EUBACT.	Mycoplasma Capric.	DQ1140	1		AGU		Plasmodium Falcip.	DS7501	6				Schizosaccha.Pom.	DM7640	8				Schizosaccha. Pom.	DR7641
2				Mycoplasma Gen.	DQ1150	2				Dictyostelium Dis.	DS7570							9				Toxoplasma Gondoii	DR7730
3				Mycoplasma Pneumo.	DQ1200	3				Podospora Anserina	DS7620	1	Ile[7]	UAG	ARCHAEA	Archaeglobus Fulg.	DI0340						
4				Acholeplasma Laid.	DQ1230	4				Podospora Anserina	DS7621	2				Methanococcus Jan.	DI0650	1	Lys[10]	UUU	ARCHAEA	Archaeglobus Fulg.	DK0340
5				Treponema Pallidum	DQ1271	5				Saccharomyces Cer.	DS7633	3				Haloferax Volcanii	RI0500	2				Methanococcus Jan.	DK0650

Table A3. (continued)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.
6				Borrelia Burgdorf.	DQ1280
7				Staphylococ. Aure.	DQ1480
8				Helicobacter Pylo.	DQ1510
9				Bacillus Subtilis	DQ1540
10				E.Coli	DQ1660
11				Haemophilus Influ.	DQ2000
12				Haemophilus Influ.	DQ2001
13				Synechocystis Sp.	DQ2140
14				Mycoplasma Capric.	RQ1140
15				E.Coli	RQ1661
1		GUC		Treponema Pallidum	DQ1270
2				Streptomyces Rim.	DQ1340
3				Streptomyces Rim.	DQ1341
4				Streptomyces Liv.	DQ1350
5				Streptomyces Liv.	DQ1351
6				E.Coli	DQ1661
7				E.Coli	RQ1660
1		GUU	EUKARYA	Plasmodium Falcip.	DQ7500
2				Trypanosoma Brucei	DQ7520
3				Trypanosoma Brucei	DQ7521
4				Leishmania Tarent.	DQ7550
5				Dictyostelium Dis.	DQ7570
6				Saccharomyces Cer.	DQ7630
7				Saccharomyces Cer.	DQ7632
8				Schizosaccha.Pom.	DQ7640
9				Toxoplasma Gondoii	DQ7730
10				Tetrahymena Therm.	RQ7542
1		GUC		Saccharomyces Cer.	DQ7631
2				Crithidia Fascic.	DQ7670
3				Leishmania Mexica.	DQ7700
4				Leptomonas Collos.	DQ7710
5				Leptomonas Seymou.	DQ7720
1	Ala[2]	CGU	ARCHAEA	Halorubrum Distri.	DA0310

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.
6				Schizosaccha. Pom.	DS7640
7				Schizosaccha. Pom.	DS7641
8				Candida Cylindra.	RS7661
1		AGC		Saccharomyces Cer.	DS7632
2				Saccharomyces Cer.	DS7634
3				Schizosaccha. Pom.	DS7642
4				Candida Cylindra.	RS7663
1	Cys[5]	ACG	ARCHAEA	Archaeglobus Fulg.	DC0340
2				Halobacterium Cut.	DC0380
3				Haloferax Volcanii	DC0500
4				Methanococcus Jan.	DC0650
1			EUBACT.	Mycoplasma Capric.	DC1140
2				Mycoplasma Gen.	DC1150
3				Mycoplasma Pneumo.	DC1200
4				Spiroplasma Melif.	DC1260
5				Treponema Pallidum	DC1270
6				Streptomyces Liv.	DC1350
7				Staphylococ. Aure.	DC1480
8				Helicobacter Pylo.	DC1510
9				Bacillus Subtilis	DC1540
10				E.Coli	DC1660
11				Haemophilus Influ.	DC2000
12				Synechocystis Sp.	DC2140
1			EUKARYA	Plasmodium Falcip.	DC7500
2				Saccharomyces Cer.	DC7630
3				Schizosaccha. Pom.	DC7640
1	Gly[5]	CCU	ARCHAEA	Archaeglobus Fulg.	DG0340
2				Methanococcus Jan.	DG0650
3				Haloferax Volcanii	RG0503
1		CCG		Halobacterium Cut.	RG0380
2				Methanococcus Jan.	DG0651
3				Haloferax Volcanii	RG0501

	Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
1		UAC		Methanococ. Vani.	DI0660
2				Methanotherm. Fer.	DI0680
1		UAU		Haloferax Volcanii	RI0501
1		UAG	EUBACT.	Bartonella Bacil.	DI1100
2				Bartonella Elizab.	DI1110
3				Bartonella Hensela	DI1120
4				Bartonella Quint.	DI1130
5				Mycoplasma Capric.	DI1141
6				Mycoplasma Gen.	DI1150
7				Acetobacter Aceti	DI1160
8				Acetobacter Europ.	DI1170
9				Acetobacter Hanse.	DI1190
10				Mycoplasma Pneumo.	DI1201
11				Acetobacter Lique.	DI1210
12				Acetobacter Lique.	DI1211
13				Acholeplasma Laid.	DI1230
14				Acetobacter Xylin.	DI1240
15				Treponema Pallidum	DI1270
16				Borrelia Burgdorf.	DI1280
17				Borrelia Burgdorf.	DI1280
18				Burkholderia Cepa.	DI1320
19				Coxiella Burnetii	DI1330
20				Gluconobacter Oxy.	DI1370
21				Lactobac.Bulgaric.	DI1500
22				Helicobacter Pylo.	DI1510
23				Lactococcus Lactis	DI1530
24				Bacillus Subtilis	DI1540
25				Bacillus Subtilis	DI1541
26				Lactobac.Acidophi.	DI1550
27				Lactobac.Casei	DI1590
28				Rhodothermus Mar.	DI1600
29				Lactobac.Curvatus	DI1610
30				Thiobacillus Ferro	DI1620
31				Lactobac.Helvetic.	DI1640

	Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
3				Methanococ.Vani.	DK0660
4				Methanotherm. Fer.	DK0680
5				Methanococ.Voltae	DK0740
6				Methanopyrus Kand.	DK0760
7				Haloferax Volcanii	RK0500
1		UUC		Archaeglobus Fulg.	DK0341
2				Haloferax Volcanii	RK0501
1		UUU	EUBACT.	Mycoplasma Capric.	DK1140
2				Mycoplasma Gen.	DK1150
3				Mycoplasma Pneumo.	DK1200
4				Mycoplasma Pg50	DK1220
5				Acholeplasma Laid.	DK1231
6				Treponema Pallidum	DK1271
7				Borrelia Burgdorf.	DK1280
8				Staphylococ. Aure.	DK1480
9				Helicobacter Pylo.	DK1510
10				Bacillus Subtilis	DK1540
11				Bacillus Subtilis	DK1541
12				E.Coli	DK1660
13				Azospirillum Lipo.	DK1720
14				Haemophilus Influ.	DK2000
15				Haemophilus Influ.	DK2001
16				Synechocystis Sp.	DK2140
17				Mycoplasma Capric.	RK1141
1		UUC		Mycoplasma Capric.	DK1141
2				Mycoplasma Gen.	DK1151
3				Mycoplasma Pneumo.	DK1201
4				Acholeplasma Laid.	DK1230
5				Treponema Pallidum	DK1270
6				Borrelia Burgdorf.	DK1281
7				Streptomyces Liv.	DK1350
8				Haemophilus Influ.	DK2002
9				Mycoplasma Capric.	RK1140
1		UUU	EUKARYA	Plasmodium Falcip.	DK7500

Table A3. (continued)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
2				Halorubrum Lacusp.	DA0320	4				Haloferax Volcanii	RG0502	32				E.Coli	DI1660	2				Trypanosoma Brucei	DK7521
3				Halorubrum Saccha.	DA0330	5				Methanobac.Therm.	RG0620	33				Mycobact.Lepra e	DI1710	3				Leishmania Tarent.	DK7550
4				Archaeglobus Fulg.	DA0340	1		CCC		Archaeglobus Fulg.	DG0341	34				Trichodesmium Sp.	DI1730	4				Dictyostelium Dis.	DK7570
5				Halorubrum Sodome.	DA0350	2				Sulfolobus Solfa.	DG0860	35				Mycoplasma Sp.	DI1760	5				Saccharomyces Cer.	DK7630
6				Halorubrum Vacuol.	DA0360	3				Thermofil. Pendens	DG0960	36				Phytoplasma Sp.	DI1770	6				Saccharomyces Cer.	RK7631
7				Natronobac. Grego.	DA0370	4				Haloferax Volcanii	RG0500	37				Aeromonas Hydroph.	DI1780	1		UUC		Trypanosoma Brucei	DK7520
8				Halobacterium Cut.	DA0380	1		CCU	EUBACT.	Mycoplasma Capric.	DG1140	38				Prevotella Rumini.	DI1790	2				Trypanosoma Brucei	DK7522
9				Natronobac. Phara.	DA0390	2				Mycoplasma Gen.	DG1151	39				Pseudomonas Cepac.	DI1810	3				Dictyostelium Dis.	DK7571
10				Halobacterium Hal.	DA0420	3				Mycoplasma Mycoid.	DG1180	40				Pseudomonas Aer.	DI1820	4				Saccharomyces Cer.	DK7631
11				Methanobac.Formi.	DA0580	4				Mycoplasma Pneumo.	DG1200	41				Pseudomonas Glad.	DI1830	5				Saccharomyces Cer.	DK7632
12				Methanobac.Therm.	DA0620	5				Treponema Pallidum	DG1270	42				Pseudomonas Fluor.	DI1840	6				Schizosaccha. Pom.	DK7640
13				Methanococcus Jan.	DA0650	6				Borrelia Burgdorf.	DG1280	43				Pseudomonas Mallei	DI1850	7				Saccharomyces Cer.	RK7630
14				Methanococ.Vani.	DA0660	7				Streptomyces Liv.	DG1351	44				Campylobac. Jejuni	DI1860						
15				Methanothrix Soeh.	DA0670	8				Staphylococ. Aure.	DG1481	45				Pseudomonas Mend.	DI1880	1	Phe^{1}_{1}	AAG	ARCHAEA	Archaeglobus Fulg.	DF0340
16				Methanotherm. Fer.	DA0680	9				Staphylococ. Aure.	DG1482	46				Caulobacter Cres.	DI1890	2				Methanococcus Jan.	DF0650
17				Methanospir. Hung.	DA0780	10				Staphylococ. Aure.	DG1483	47				Brucella Suis	DI1900	3				Methanococ.Vani.	DF0660
18				Thermococcus Celer	DA0940	11				Helicobacter Pylo.	DG1511	48				Brucella Melitens.	DI1910	4				Sulfolobus Solfa.	DF0860
19				Thermoprot. Tenax	DA0980	12				Lactococcus Lactis	DG1530	49				Brucella Abortus	DI1920	5				Haloferax Volcanii	RF0500
20				Haloferax Volcanii	RA0502	13				Bacillus Subtilis	DG1540	50				Brucella Abortus	DI1921	1			EUBACT.	Mycoplasma Capric.	DF1140
1		CGG		Archaeglobus Fulg.	DA0342	14				Bacillus Subtilis	DG1541	51				Ochrobactrum Anth.	DI1960	2				Mycoplasma Gen.	DF1150
2				Methanococcus Jan.	DA0651	15				Stigmatella Aurant	DG1630	52				Pseudomonas Pick.	DI1970	3				Mycoplasma Mycoid.	DF1180
3				Haloferax Volcanii	RA0501	16				E.Coli	DG1660	53				Pseudomonas Pseud.	DI1990	4				Acholeplasma Laid.	DF1230
1		CGC		Archaeglobus Fulg.	DA0341	17				Pseudomonas Aer.	DG1820	54				Haemophilus Influ.	DI2001	5				Spiroplasma Melif.	DF1260
2				Thermoprot. Tenax	DA0981	18				Campylobac.Jejuni	DG1860	55				Salmonella Enteri.	DI2040	6				Treponema Pallidum	DF1270
3				Halobacterium Cut.	RA0380	19				Rickettsia Prow.	DG1870	56				Stenotro.Maltoph.	DI2080	7				Borrelia Burgdorf.	DF1280
4				Haloferax Volcanii	RA0500	20				Haemophilus Influ.	DG2002	57				Xanthomonas Campe.	DI2090	8				Borrelia Burgdorf.	DF1281
1		CGU	EUBACT.	Bartonella Elizab.	DA1110	21				Synechocystis Sp.	DG2140	58				Anacystis Nidulans	DI2100	9				Staphylococ. Aure.	DF1480
2				Bartonella Quint.	DA1130	22				Staphylococ. Epid.	RG1380	59				Synechocystis Sp.	DI2141	10				Helicobacter Pylo.	DF1510
3				Mycoplasma Capric.	DA1140	23				Staphylococ. Epid.	RG1381	60				Synechocystis Sp.	DI2142	11				Lactococcus Lactis	DF1530
4				Mycoplasma Gen.	DA1150	24				E.Coli	RG1662	61				Mycoplasma Mycoid.	RI1180	12				Bacillus Subtilis	DF1540
5				Acetobacter Aceti	DA1160	25				Salmonella Typhi.	RG1701	62				Thermus Thermophi.	RI1580	13				Bacillus Subtilis	DF1541
6				Acetobacter Europ.	DA1170	1		CCG		Mycoplasma Gen.	DG1150	63				E.Coli	RI1661	14				E.Coli	DF1660
7				Mycoplasma Mycoid.	DA1180	2				Mycoplasma Pneumo.	DG1201	1		UAC		Mycoplasma Capric.	DI1140	15				Haemophilus Influ.	DF2000
8				Acetobacter Hanse.	DA1190	3				Acholeplasma Laid.	DG1230	2				Mycoplasma Gen.	DI1151	16				Haemophilus Influ.	DF2001

Table A3. (continued)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.		Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
43				Trichodesmium Sp.	DA1730	2				Methanococcus Jan.	DV0652	2				Methanococcus Jan.	DL0652	13				E.Coli	DY1660
44				Aeromonas Hydroph.	DA1780	3				Sulfolobus Solfa.	DV0860	3				Haloferax Volcanii	RL0503	14				E.Coli	DY1661
45				Prevotella Rumini.	DA1790	4				Halobacterium Cut.	RV0382	1		GAU	EUBACT.	Mycoplasma Capric.	DL1141	15				Pseudomonas Aer.	DY1820
46				Pseudomonas Cepac.	DA1810	5				Haloferax Volcanii	RV0501	2				Mycoplasma Gen.	DL1151	16				Campylobac.Jejuni	DY1860
47				Pseudomonas Aer.	DA1820	1		CAC		Archaeglobus Fulg.	DV0340	3				Mycoplasma Pneumo.	DL1201	17				Rickettsia Prow.	DY1870
48				Pseudomonas Glad.	DA1830	2				Methanococcus Jan.	DV0650	4				Mycoplasma Pg50	DL1220	18				Haemophilus Influ.	DY2000
49				Pseudomonas Fluor.	DA1840	3				Halobacterium Cut.	RV0380	5				Acholeplasma Laid.	DL1231	19				Synechocystis Sp.	DY2140
50				Pseudomonas Mallei	DA1850	4				Halobacterium Cut.	RV0381	6				Treponema Pallidum	DL1273	20				Bacillus Stearo.	RY2120
51				Campylobac.Jejuni	DA1860	5				Haloferax Volcanii	RV0500	7				Borrelia Burgdorf.	DL1282	1			EUKARYA	Plasmodium Falcip.	DY750O
52				Pseudomonas Mend.	DA1880	1		CAU	EUBACT.	Mycoplasma Capric.	DV1140	8				Staphylococ. Aure.	DL1480	2				Trypanosoma Brucei	DY7520
53				Caulobacter Cres.	DA1890	2				Mycoplasma Mycoid.	DV1180	9				Helicobacter Pylo.	DL1513	3				Tetrahymena Therm.	DY7540
54				Brucella Suis	DA1900	3				Mycoplasma Pneumo.	DV1200	10				Bacillus Subtilis	DL1544	4				Dictyostelium Dis.	DY7570
55				Brucella Melitens.	DA1910	4				Acholeplasma Laid.	DV1230	11				E.Coli	DL1661	5				Saccharomyces Cer.	DY7630
56				Brucella Abortus	DA1920	5				Treponema Pallidum	DV1271	12				Photobac. Leiogna.	DL1750	6				Saccharomyces Cer.	DY7631
57				Brucella Abortus	DA1921	6				Borrelia Burgdorf.	DV1280	13				Haemophilus Influ.	DL2004	7				Leishmania Donava.	DY7690
58				Ochrobactrum Anth.	DA1960	7				Mycobact. Tuberc.	DV1400	14				Synechocystis Sp.	DL2140	8				Scenedesmus Obliq.	RY7560
59				Pseudomonas Pick.	DA1970	8				Staphylococ. Aure.	DV1480	15				Mycoplasma Capric.	RL1142	9				Schizosaccha. Pom.	RY7640
60				Pseudomonas Pseud.	DA1990	9				Lactobac.Bulgaric.	DV1500	1		GAG		Mycoplasma Gen.	DL1152	10				Torulopsis Utilis	RY7650
61				Haemophilus Influ.	DA2001	10				Helicobacter Pylo.	DV1511	2				Treponema Pallidum	DL1271						
62				Stenotro.Maltoph.	DA2080	11				Bacillus Subtilis	DV1540	3				Borrelia Burgdorf.	DL1280	1	His[13]	GUG	ARCHAEA	Archaeglobus Fulg.	DH0340
63				Anacystis Nidulans	DA2100	12				Bacillus Sp. Ps3	DV1570	4				Helicobacter Pylo.	DL1512	2				Methanococcus Jan.	DH0650
64				Synechocystis Sp.	DA2142	13				E.Coli	DV1660	5				Bacillus Subtilis	DL1545	3				Methanococ.Vani.	DH0660
65				Streptococcus Sal.	DA2160	14				Azospirillum Lipo.	DV1720	6				E.Coli	DL1663	4				Methanotherm. Fer.	DH0680
66				Streptococcus Pn.	DA2170	15				Haemophilus Influ.	DV2000	7				Haemophilus Influ.	DL2002	5				Halobacterium Cut.	RH0380
67				E.Coli	RA1661	16				Synechocystis Sp.	DV2140	8				Synechocystis Sp.	DL2141	6				Haloferax Volcanii	RH0500
68				E.Coli	RA1662	17				Synechococcus Sp.	DV2150	9				E.Coli	RL1662	1			EUBACT.	Mycoplasma Capric.	DH1140
1		CGG		Treponema Pallidum	DA1271	18				E.Coli	RV1662	1		GAC		Treponema Pallidum	DL1274	2				Mycoplasma Gen.	DH1150
2				Helicobacter Pylo.	DA1510	1		CAG		Treponema Pallidum	DV1270	2				Bacillus Subtilis	DL1540	3				Mycoplasma Pneumo.	DH1200
3				E.Coli	DA1661	2				Streptomyces Liv.	DV1350	3				E.Coli	DL1660	4				Acholeplasma Laid.	DH1230
4				Haemophilus Influ.	DA2000	3				Streptomyces Liv.	DV1351	4				Salmonella Typhi.	DL1700	5				Treponema Pallidum	DH1270
5				Synechocystis Sp.	DA2141	4				Helicobacter Pylo.	DV1510	5				Mycobact.Leprae	DL1710	6				Borrelia Burgdorf.	DH1280
6				E.Coli	RA1660	5				E.Coli	DV1661	6				Aeromonas Hydroph.	DL1780	7				Staphylococ. Aure.	DH1480
1		CGC		Treponema Pallidum	DA1270	6				E.Coli	DV1662	7				Rhizobium Meliloti	DL1940	8				Helicobacter Pylo.	DH1510
2				Synechocystis Sp.	DA2140	7				Haemophilus Influ.	DV2001	8				Bordetella Pertus.	DL1980	9				Bacillus Subtilis	DH1540

Table S3. (continued)

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.
1		CGU	EUKARYA	Saccharomyces Cer.	DA7630
2				Plasmodium Falcip.	DA7500
3				Toxoplasma Gondoii	DA7730

266

	Amino acid	Anti codon 3'-5'	Domain	Species	Access no.
8				Synechocystis Sp.	DV2141
9				E.Coli	RV1660
10				Bacillus Stearo.	RV2120

266

	Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
9				Synechocystis Sp.	DL2144
10				E.Coli	RL1661
11				Anacystis Nidulans	RL2101

266

	Amino acid	Anti-codon 3'-5'	Domain	Species	Access no.
10				Bacillus Subtilis	DH1541
11				E.Coli	DH1660
12				Salmonella Typhi.	DH1700

265

= 1063

	Amino acid	Anti-codon 3'- 5'	Domain	Species	Access no.
13	His[13]	GUG	EUBACT.	Photobact. Phosph.	DH1740
14				Aeromonas Hydroph.	DH1780
15				Haemophilus Influ.	DH2000
16				Synechocystis Sp.	DH2140
1			EUKARYA	Plasmodium Falcip.	DH7500
2				Dictyostelium Dis.	DH7570
3				Saccharomyces Cer.	DH7630
4				Schizosaccha.Pom.	DH7640
5				Saccharomyces Cer.	RH7630
6				Saccharomyces Cer.	RH7631
1	Trp[14]	ACC	ARCHAEA	Archaeglobus Fulg.	DW0340
2				Halobacterium Med.	DW0460
3				Haloferax Volcanii	DW0500
4				Methanococcus Jan.	DW0650
1			EUBACT.	Thermotoga Marit.	DW0990
2				Mycoplasma Capric.	DW1141
3				Mycoplasma Gen.	DW1150
4				Mycoplasma Pneumo.	DW1200
5				Acholeplasma Laid.	DW1230

19

	Amino acid	Anti-codon 3'- 5'	Domain	Species	Access no.
6	His[13]	GUG		Spiroplasma Citri	DW1251
7				Treponema Pallidum	DW1270
8				Borrelia Burgdorf.	DW1280
9				Streptomyces Gris.	DW1300
10				Staphylococ. Aure.	DW1480
11				Helicobacter Pylo.	DW1510
12				Bacillus Subtilis	DW1540
13				E.Coli	DW1660
14				Rickettsia Prow.	DW1870
15				Haemophilus Influ.	DW2000
16				Synechocystis Sp.	DW2140
1			EUKARYA	Plasmodium Falcip.	DW7500
2				Leishmania Tarent.	DW7550
3				Dictyostelium Dis.	DW7570
4				Saccharomyces Cer.	DW7630
5				Saccharomyces Cer.	DW7631
6				Schizosaccha.Pom.	DW7640
7				Toxoplasma Gondoii	DW7730

18

= 37 + 1063 = 1100

Table A3. (continued)

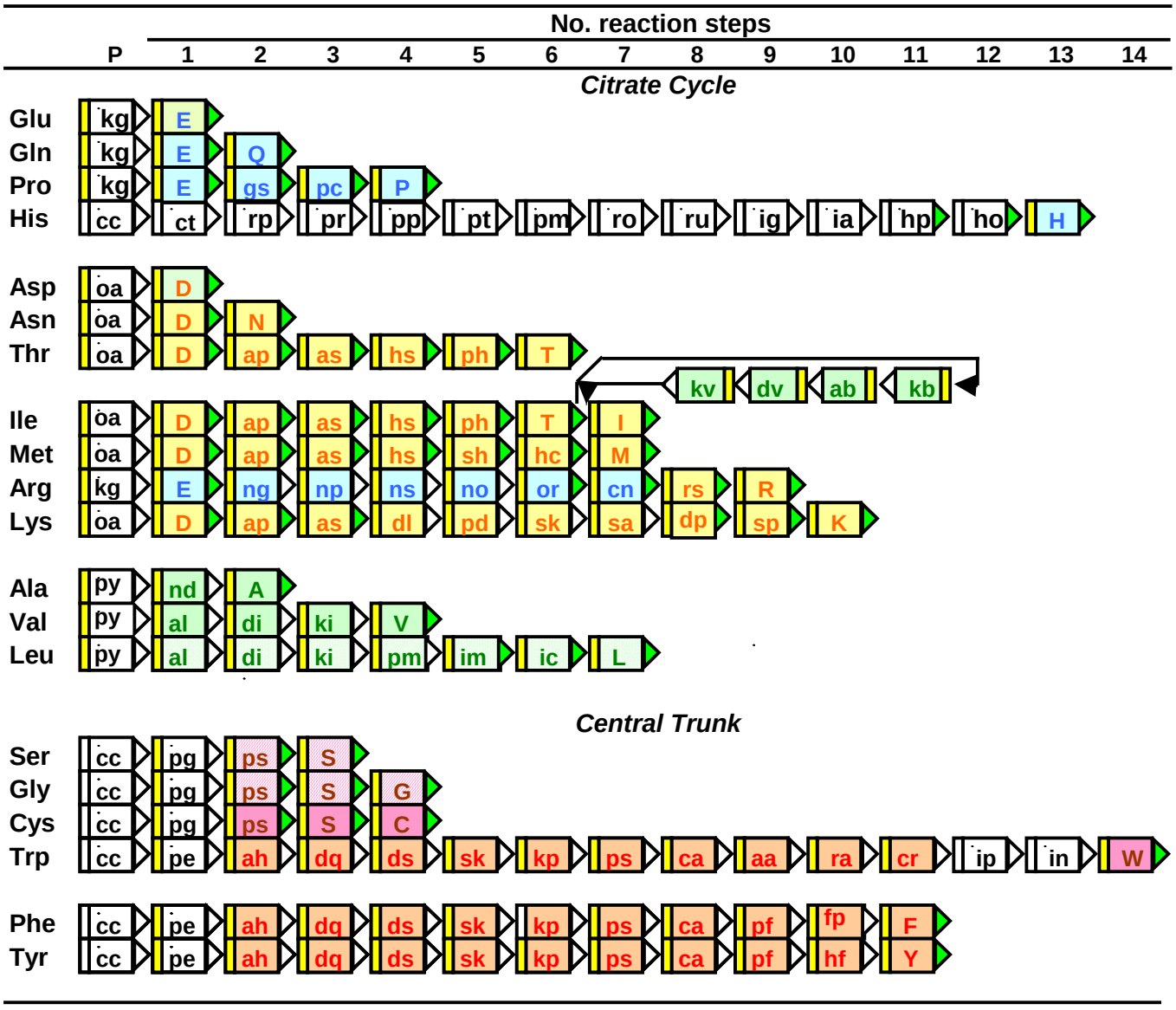


Figure. A1. Synthesis pathways of amino acids in the standard code. The number of reaction steps in each path appears in the overbar. Thirteen amino acid pathways originate from branch reactions at oxaloacetate (oa), a-ketoglutarate (kg), and pyruvate (py) in the citrate cycle (cc) . Six originate from central trunk (ct) precursors (P), 3-phosphoglycerate (pg) and phosphoenolpyruvate (pe) (with erythrose-4-phosphate from pentose cycle). One amino acid (histidine) arose from ribose-5-phosphate (rp) in the pentose cycle. Gold bar (left side) indentifies intermediates with an α-carboxyl, the tRNA attachment site. All intermediates contain a free α-carboxyl, except in the histidine pathway and step-12 and -13 in tryptophan synthesis. Green triangle (right side) denotes an α-amine; generally this peptide-bond forming group is acquired near the path end, except for methionine and

related aa. Isoleucine path incorporates four (uncounted) reaction steps from the valine pathway. Three letter amino acid abbreviations appear in left-hand column. Upper-case, single letter amino acid abbreviations are used within pathways. Letter and background colors of intermediates correspond to those in Fig. 2 specifying code domains and quasidomains. Lower-case, double letter abbreviations identify non-amino acid intermediates[31]**:** gs, glutamate-γ-semialdehyde; pc, Δ1-pyrroline-5'-carboxylate (**Pro**). pp, phosphatidyl-ribosyl-pyrophosphate; pt, phospho-ribosyl-adenosine-triphosphate; pm, phospho-ribosyl-adenosine-monophosphate; ro, phospho-ribosyl-formimino-amino-imidazole-carboxamide-ribose-phosphate; ru, phospho-ribulosyl-formimino-amino-imidazole-carboxamide-ribose-phosphate; ig, erythro-imidazole-gylcerol-phosphate; ia, imidazole-acetol-phosphate; hp, histidinol-phosphate; ho, histidinol (**His**). ap, aspartyl-phosphate; as, aspartate-β-semialdehyde; hs, homoserine; ph, o-phospho-homoserine (**Thr**). kb, a-keto-butyrate; ab, a-aceto-a-hydroxy-butyrate; dv, a,β-dihydroxy-isovalerate; kv, a-keto-isovalerate (**Ile**). ng, N-acetyl-glutamate; np, N-acetyl-glutamate-phosphate; ns, N-acetyl-glutamate-γ -semialdehyde; no, N-acetyl-ornithine; or, ornithine; cn, citrulline; rs, arginine-succinate (**Arg**). dl, a,β-dihydropicolineate; pd, Δ^1-piperdiene-2,6-dicarboxylate; sk, N-succinyl-ε-keto-a-amino-pimelate; sa, N-succinyl-a,ε-diamino-pimelate; dp, a,ε-diamino-L-pimelate; sp, meso-aε-diamino-pimelate (**Lys**). nd, Glu amine-donor (**Ala**). al, a-aceto-lactate; dl, a,β-dihydroxy-isovalerate;kl, a-keto-isovalerate (**Val**). pm, a-isopropyl-malate; im, β-isopropyl-malate; ic, a-keto-isocaproate (**Leu**). ah, β-deoxy-arabino-heptulosonate-7-phosohate; dq, 5-dehydroquinate; ds, 5-dehydro-shikimate; sk, shikimate; kp, shikimate-5-phisohate; ps, 3-enolpyruvyl-shikimate-5-phosphate; ca, chorismate; aa, anthranilate; ra, N-phospho-ribosyl-anthranilate; cr, 1-(o-carboxyphenylamino)-1'-deoxyribulose-5-phosphate; ip, indole-3-glycerol-phosphate; in, indole (**Trp**). pf, prephenate; fp, phenyl-pyruvate (**Phe**). hf, p-hydroxy-phenyl-pyruvate (**Tyr**).

References

1. Crick FHC (1966) Genetic code – yesterday, today, and tomorrow. Cold Spring Harbor Symp. Quant. Biol. 31: 1-5.
2. Davis BK (2007) Making sense of the genetic code with the path-distance model. In: Ostrovisky MH, editor. Leading-edge messenger RNA research communications. New York: Nova Science. pp. 1-32.
3. Crick FHC (1968) The origin of the genetic code. J. Mol. Biol. 38: 367-379.

4. Davis BK (2008) Deep-structure of the ‘universal’ genetic code and the origin of proteins. FEBS J. 275 (Supp. 1): 75.

5. Davis BK (1999) Evolution of the genetic code. Prog. Biophys. Mol. Biol. 72: 157-243.

6. Davis BK (2002) Molecular evolution before the origin of species. Prog. Biophys. Mol. Biol. 79: 77-133.

7. Davis BK (2008) Imprint of early tRNA diversification on the genetic code: Domains of contiguous codons read by related adaptors for sibling amino acids. In: Takayama, T, editor. Messenger RNA research perspectives. New York: Nova Science. pp. 35-79

8. Woese CR (1965) Order in the genetic code. Proc. Natl. Acad. Sci. U.S.A. 54: 71-75.

9. Freeland SJ, Hurst L (1998) Load minimization of the genetic code: history does not explain the pattern. Proc. R. Soc. Lond. B 265: 2111-2119.

10. Taylor FJR, Coates D (1989) The code within codes. Biosystems 22: 177-187.

11. Rodin AS, Szathmary E, Rodin SN (2009) One ancestor for two codes viewed from the perspective of two complementary modes of tRNA aminoacylation. Biology Direct 4-4: 1-30.

12. McClain W H (1994) Rules that govern tRNA identity on protein synthesis. J. Mol. Biol. 234: 257-280.

13. Saks ME, Sampson J R, Abelson JN (1994) The transfer RNA identity problem: a search for rules. Science 263: 191-197.

14. Giege R (2008) Toward a more complete view of tRNA biology. Nature Struct. Mol. Biol. 15: 1007-1014.

15. DeDuve C (1988) The second genetic code. Nature 333: 117-118.

16. Lim V, Curran P (2001) Analysis of codon:anticodon interactions within the ribosome provides new insights into code reading and genetic code structure. RNA 7: 942-957.

17. Walter KU, Vamvaca K, Hilvert D (2005) An active enzyme constructed from a 9-amino acid alphabet. J. Biol. Chem. 280: 37742-37746.

18. Griffiths G (2007) Cell evolution and the problem of membrane topology. Nature Reviews:

Mol. Cell Biol. 8: 1018-1024.

19. Davis BK (2009) On mapping the genetic code. J. Theor. Biol. 360: 860-862.
20. Eigen M, Lindemann BF, Tietz M, Winkler-Oswatitsch R, Dress A et al. (1989) How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 244: 673-679.
21. Saks ME, Sampson JR (1995) Evolution of tRNA recognition systems and tRNA gene sequences. J. Mol. Evol. 40: 509-518.
22. Danchin A (1989) Homeotopic transformation and the origin of translation. Prog. Biophys. Mol. Biol. 54: 81-86.
23. Soll D (1993) Transfer RNA: An RNA for all seasons. In: Gesteland RF, Atkins JF, editors. The RNA World. Plainview: Cold Spring Harbor Laboratory Press. pp. 157-184.
24. Wilcox M, Nirenberg M (1968) Transfer RNA as a cofactor coupling aa synthesis with that of protein. Proc. Natl. Acad. Sci. U.S.A. 55: 229-236.
25. Schon A, Kannangara CG, Gough S, Soll D (1988) Protein biosynthesis in organelles requires misacylation of tRNA. Nature 331: 187-190.
26. Marcker K, Sanger F (1964) N-Formyl-methionyl-sRNA. J. Mol. Biol. 8: 835-840.
27. Sauerwald A, Zhu W, Major TA, Roy H, Palioura S et al. (2005) RNA-dependent cysteine biosynthesis in archaea. Science 307: 1969-1972.
28. Forchammer K, Bock A (1991) Selenocysteine synthesis from Escherichia coli – analysis of the reaction sequence. J. Biol. Chem. 266: 6324-6328.
29. Kinchin AI (1957) Mathematical Foundations of Information Theory. New York: Dover.
30. Marck C, Grosjean H (2002) tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. RNA 8: 1189-1232.
31. Michal G (1978) Biochemical Pathways. Indianopolis: Boehringer Mannheim Biochemicals. 3rd Printing.
32. Efron B (1992) Six questions raised by the bootstrap. In: LePage R, Billard L, editors. Exploring

the Limits of the Bootstrap. New York: Wiley-Interscience. pp. 99-126.

33. Fisher RA (1958) Statistical Methods for Research Workers. Edinburgh: Oliver & Boyd.
34. Bulmer MG (1979) Principles of Statistics. New York: Dover.
35. Allwood AC, Walter MR, Burch IW, Kamber BS (2007) 3.43 billion-year-old stromatolite reef from the Pilbara Craton of Western Australia**:** Ecosystem-scale insights to early life on Earth. Precambrian Res. 158: 198-227.
36. McGuinness E (2010) Some molecular moments of the Hadean and Archaean aeons: retrospective overview from the interfacing years of the second and third millennia. Chem. Rev. 110: 5191-5215.
37. Davis BK (1998) The forces driving molecular evolution. Prog. Biophys. Mol. Biol. 69: 83-150.
38. Shepard K, et al. (2008) From one amino acid to another**:** tRNA-dependent aa biosynthesis. Nuc. Acids Res. 36: 1813-1825.
39. Migita LK, Doi RH (1970) Formylation of methionyl-transfer RNA from prokaryotes and eukaryotes by Bacillus subtilis transformylase. Arch. Biochem. Biophys. 138, 457-463.
40. Francklyn C, Perona JJ, Puetz J, Hou Y-A (2002) Aminoacyl-tRNA stnthetases: Versatile players in the changing theater of translation. RNA 8, 1363-1372.
41. Doudna JA, Lorsch JR (2005) Ribozyme catalysis: not different, just worse. Nature: Struc. Mol. Biol. 12: 395-402.
42. Nirenberg M, Caskey T, Marshall R, Brimbacombe R, Kellogg D et al. (1966) The RNA code and protein synthesis. Cold Spring Harbor Symp. Quant. Biol. 31: 11-24.
43. Dillon LS (1973) The origins of the genetic code. Botanical Rev. 39: 301-345.
44. Dunnill P (1966) Triplet-nucleotide-amino-acid pairing: a stereochemical basis for the division between protein and non-protein amino acids. Nature 210: 1267-1268.
45. Wachterhauser G. (1992) Groundworks for an evolutionary biochemistry: the iron-sulphur world. Prog. Biophys. Mol. Biol. 58: 85-201.
46. Nitschke W, Russell MJ (2009) Hydrothermal focusing and chemical and chemiosmotic energy,

supported by delivery of catalytic Fe, Ni, Mo/W, Co, S and Se, forced life to emerge. J. Mol. Evol. 69: 481-496.

47. Wong JT-F (1975) A coevolution theory of the genetic code. Proc. Natl. Acad. Sci U.S.A. 72: 1909-1912.

48. Perlwitz MD, Burks C, Waterman MS (1988) Pattern analysis of the genetic code. Advan. App. Math. 9: 7-21.

49. Garrett RH, Grisham CM (1999) Biochemistry. San Diego: Saunders.

50. Brooks DJ, Fresco JR, Lesk AM, Singh M (2002) Evolution of amino acid frequencies in proteins over deep time: Inferred order of introduction of amino acids into the genetic code. Mol. Biol. Evol. 19: 1645-1655.

51. Biro JC, Benyo B, Sansom C, Szlavecz A, Fordos G et al. (2003) A common periodic table of codons and amino acids. Biochem. Biophys. Res. Comm. 306: 408-415.

52. Jukes TH (1973) Arginine as an evolutionary intruder into protein synhesis. Biochem. Biophys. Res Commun. 53: 709-714.

53. Brooks DJ, Fresco JR (2003) Greater GNN pattern bias in sequence elements encoding conserved residues of ancient proteins may be an indicator of amino acid composition of early proteins. Gene 303: 177-185.

54. Sonneborn TM (1965) Degeneracy of the genetic code: extent, nature, and genetic implications. In: Bryson V, Vogel HJ, editors. Evolving Genes and Proteins. New York: Academic Press. pp. 379-397

55. Hinegardner RT, Engelberger J (1963) Rationale for a universal genetic code. Science 142: 1083-1085.

56. Davis BK (2004) Expansion of the genetic code in yeast: making life more complex. BioEssays 26: 111-115.

57. Deamer DW (2008) How leaky were primitive cells? Nature 454: 37-38.

58. Christopherson S (2004).Reconstruction and preliminary characterization of the evolutionary

origin of iron-sulfur proteins: the oldest k known protein and its relation to the origin of life. Thesis **s991053.** Lyngby: Technical University of.Denmark,

59. Maizel N, Weiner AM (1994) Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication not translation. Proc. Natl. Acad. Sci. U.S.A. 91: 6729-6734.
60. Tlusty T (2010) A colorful origin for the genetic code: information theory, statistical mechanics and the emergence of molecular codes. Phys. Life Rev. 7: 362-376.
61. Bretscher MS, Goodman HM, Menninger JR, Smith JD (1965) Polypeptide chain termination using synthetic polynucleotides. J. Mol. Biol. 14: 634-639.
62. Sun F-J, Caetano-Anolles G (2008) Evolutionary patterns in the sequence and structure of transfer RNA: a window into early translation and the genetic code. PLoS One 3: e2799.
63. Kyprides N, Overbeek R, Ouzounis C (1999) Universal protein families and the functional content of the last universal common ancestor. J. Mol. Evol. 49: 413-423.
64. Miller SL,.Orgel L (1974) The Origin of Life on the Earth. Engelwood-Cliffs: Prentice-Hall.
65. Trifonov EN (2000) Consensus temporal order of amino acids and evolution of the triplet code. Gene 261:139-151. .
66. Woese CR, Olsen GJ, Ibba M, Soll D (2000).Aminoacyl-tRNA synthetases, the genetic code and the evolutionary process. Microbiol. Mol. Biol. Rev. 64: 202-236.
67. Yarus M (2000) RNA-ligand chemistry: A testable source for the genetic code. RNA 6: 475-484.
68. Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok, M (1996).Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. Science 272:1343-1347.
69. Davis BK (2008) Comments on the search for the source of the genetic code. In: Takeyama T, editor. Messenger RNA research perspectives. New York: Nova Science. pp. 1-8.